

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 08:18:22 ; Search time 10398.6 Seconds

(without alignments)
10745.552 Million cell updates/sec

Title: US-10-603-260-1

Perfect score: 2578

Sequence: 1 agcttcgacattacagcca.....tgataataatgcttcttag 2578

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

Listing first 45 summaries

```

1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vt:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rod:*
36: em_hg_mam:*
37: em_hg_vit:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|----------------------|
| C 1 | 574 | 22.3 | 10977 | 1 | AE004352 | AE004352 Vibrio ch |
| C 2 | 467 | 18.1 | 306147 | 1 | AE016805 | AE016805 Vibrio vu |
| C 3 | 462.2 | 17.9 | 241900 | 1 | AP005337 | AP005337 Vibrio vu |
| C 4 | 403 | 15.6 | 303450 | 1 | AP005085 | AP005085 Vibrio pa |
| C 5 | 230.4 | 8.9 | 247450 | 1 | AP005349 | AP005349 Vibrio vu |
| C 6 | 230.4 | 8.9 | 302331 | 1 | AE016810 | AE016810 Vibrio vu |
| C 7 | 226.4 | 8.8 | 4928 | 6 | AX023555 | AX023555 Sequence |
| C 8 | 226.4 | 8.8 | 5391 | 6 | AX023554 | AX023554 Sequence |
| C 9 | 226.4 | 8.8 | 7057 | 6 | AX023552 | AX023552 Sequence |
| C 10 | 226.4 | 8.8 | 9402 | 6 | AX023551 | AX023551 Sequence |
| C 11 | 226.4 | 8.8 | 10890 | 6 | AX023553 | AX023553 Sequence |
| C 12 | 224.8 | 8.7 | 1073 | 6 | AR371489 | AR371489 Sequence |
| C 13 | 224.8 | 8.7 | 1073 | 6 | AX195443 | AX195443 Sequence |
| C 14 | 224.8 | 8.7 | 2297 | 6 | BD261822 | BD261822 Recombina |
| C 15 | 224.8 | 8.7 | 2297 | 6 | AX027819 | AX027819 Sequence |
| C 16 | 224.8 | 8.7 | 2368 | 12 | AF264618 | AF264618 Cloning v |
| C 17 | 224.8 | 8.7 | 2417 | 12 | CVSPSP70 | CVSPSP70 Cloning v |
| C 18 | 224.8 | 8.7 | 2419 | 12 | CVSPSP71 | CVSPSP71 Cloning vec |
| C 19 | 224.8 | 8.7 | 2450 | 12 | BD002331 | BD002331 Cloning vec |
| C 20 | 224.8 | 8.7 | 2462 | 6 | BD002331 | BD002331 Transgeni |
| C 21 | 224.8 | 8.7 | 2462 | 12 | CVSPSP72 | CVSPSP72 Cloning vec |
| C 22 | 224.8 | 8.7 | 2464 | 12 | CVSPSP73 | CVSPSP73 Cloning vec |
| C 23 | 224.8 | 8.7 | 2512 | 12 | ASA2679 | ASA2679 Synthetic |
| C 24 | 224.8 | 8.7 | 2580 | 6 | AR199681 | AR199681 Sequence |
| C 25 | 224.8 | 8.7 | 2655 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 26 | 224.8 | 8.7 | 2655 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 27 | 224.8 | 8.7 | 2668 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 28 | 224.8 | 8.7 | 2674 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 29 | 224.8 | 8.7 | 2680 | 6 | AO2712 | AO2712 Cloning |
| C 30 | 224.8 | 8.7 | 2680 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 31 | 224.8 | 8.7 | 2680 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 32 | 224.8 | 8.7 | 2686 | 6 | AO2710 | AO2710 Cloning |
| C 33 | 224.8 | 8.7 | 2686 | 6 | AX743996 | AX743996 Sequence |
| C 34 | 224.8 | 8.7 | 2686 | 6 | BD000457 | BD000457 Sequence |
| C 35 | 224.8 | 8.7 | 2686 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 36 | 224.8 | 8.7 | 2686 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 37 | 224.8 | 8.7 | 2695 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 38 | 224.8 | 8.7 | 2695 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 39 | 224.8 | 8.7 | 2695 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 40 | 224.8 | 8.7 | 2695 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 41 | 224.8 | 8.7 | 2695 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 42 | 224.8 | 8.7 | 2701 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 43 | 224.8 | 8.7 | 2704 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 44 | 224.8 | 8.7 | 2704 | 12 | SYNPU88V | SYNPU88V Cloning |
| C 45 | 224.8 | 8.7 | 2710 | 12 | SYNPU88V | SYNPU88V Cloning |

ALIGNMENTS

RESULT 1
LOCUS AE004352 10977 bp DNA linear BCT 10-OCT-2003
DEFINITION Vibrio cholerae O1 biovar eltor str. N16961
9 of 93 of the complete chromosome.
ACCESSION AE004352 AE003853
VERSION AE004352.1 GI:9657475
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae O1 biovar eltor str. N16961
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
1 (bases 1 to 10977)
Heidelberg, D.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.T.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,

TITLE
JOURNAL
MEDLINE
PubMed
10952301
2 (bases 1 to 10977)
Heidelberg, J.F., Bisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Emolaeva, M.D., Vamathevan, J., Bae, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Ueberlack, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1. .10977
/organism="Vibrio cholerae O1 biovar eltor str. N16961"
/mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:243277"
/chromosome="II"
/note="biotype: El Tor"
complement (81. .974)
/gene="VCA0090"
complement (81. .974)
/gene="VCA0090"
complement (81. .974)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF6004.1"
/db_xref="GI:9657476"
/translation="MSLVNLIQLCLAIWGSFLFMRFAISFGAYLIARVGFPA
LSLFLVAQLRLRSLPRQHPHFLILGINTAVPFLIFALMLTAVSLTANSTAP
IWGAVIGFWMHGTPLSRKAVAGILLVSGVAVIGDMVAIGHAALPVCALTAAS
YGLAVYTKQAPQLSAPENAGSMACLVAVPLMVPRLRFTPSLSLGAAYTLGGVI
CTGLAYIVFRLVKAIGAASTLSVTFLIFVFGILMGWYILDPGIGNTLFGTLLVAG
TMEVTGSLNARLSRQPPH"
1204. .1500
/gene="VCA0091"
1204. .1500
/gene="VCA0091"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF6005.1"
/db_xref="GI:9657477"
/translation="MARPKIPRIECPHSPKSPKNGVPRIQLARVELADELAIRL
VDIQLQQQQAALQVSRQTLNLYKAARFVVDLILQKALYIQAIDNKSDD"
complement (1671. .2300)
/gene="VCA0092"
complement (1671. .2300)
/gene="VCA0092"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF6006.1"
/db_xref="GI:9657478"
/translation="MKTVIIHGIYMGUWQPLAORLNTGYOTEVISYNTLAIDEX
VFQITSDALQDRINLVGSLGLMFKHILSRHSPPVTSVVALAPLIGASTVP
KIQQLGAMLGKNAHLYGLQHDQSWELPQRLICAGTLAGTRFPIILLGSGSGCDSTV
TVARTQISGWTDLHLHQSHTGIVSHKTRQIDYFIRHQFQHKRIP"
2400. .2573
/gene="VCA0093"
CDS
2400. .2573
/gene="VCA0093"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF6007.1"
/db_xref="GI:9657479"
/translation="MEQPCWLRSVNPITWFKYVNEDEKCLLPVTSRLAKLEKXVH
DDCLFKKSQDLM"
2801. .3157
/gene="VCA0094"
2801. .3157
/gene="VCA0094"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF6008.1"
/db_xref="GI:9657480"
/translation="WDVNIKSPRIPIINDYALQOLERISTKWTTHIMHDSBSARF
GELOYSLNGISAKRLSHRLMEIGVMWEQOPTLPVTVYGLTSGRVVANGLTRW
IDHYVENERPLESN"
3426. .4124
/gene="VCA0095"
3426. .4124
/gene="VCA0095"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF6009.1"
/db_xref="GI:9657481"
/translation="MYQMKLSMPLFSFVTLFPLGISLAFYTSSTPRNLPPBROH
VIAPKAPBETLSIHERPAVELESTERVYVNEBVOGEAPRVIPQULBQWLCI
NPTPELADLKRSQAVTDIOALEBENDQFVLMAGSATPLPNGSIRLILBENT
QAQGIQTMAGNFEFGTPTFTFGHSLMGFIGHQSGIKLEGVQAMVYQVPS
HGFEDTQSHDRAH"
4167. .5390
/gene="VCA0096"
4167. .5390
/gene="VCA0096"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF6010.1"
/db_xref="GI:9657482"
/translation="MKKASIALAVSVGLCSPALMATTEIDVLTGYTPDTAKGPKOEAV
AOWHVSNAKVKLKDGLDIKVNLAATKSVQDTPQGLKKSQSEVIDAATPRRIRDP
AFADVAVRQOVGADWVAIFRYLDVNNSPYERQPNQSYISGLANI VAPSAKTRQ
NAKSMYSHSLNECGAETFIHELGVNFGNHAHEQYRELPNNHNGEVAYGKIG
QFATIMAYPHLFGVGRSYKFSPPYLOCEGAPGVKDYANVRALGLTAPIAOYVITG
KRPDIDNPGDTEPTDNTNNVFTIKGGLAPDMKILTLPIVVSQSGSTQVAIDITH
EYRDLNSTRLPADPGSYVWLKQNRVYRQGSYVQVFLNVDPSSAAGEWRLFIQDHF
GSKLTAKQFOTTP"
6119
complement (5403. .6119)
/gene="VCA0097"
complement (5403. .6119)
/gene="VCA0097"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF6011.1"
/db_xref="GI:9657483"
/translation="MIVTIDMI CURLAPKGIETLLVYRSNPNRBDGIMALPGWVPD
EDLSAOGEPADKDPDSARRICQKITHYPTNFISSDPLVDGNPFGDSGMSISISHA
LINSNVTQIEAEMDSRYVWPALEITLQGVYALPDHVAQIOHAWKILRAAYEYIS
VTVLSIRFETPVADIIETAYAKSGDVARMYTKRLIDTGVIVSNKXLAASKGKGKGP
ATVRLANSNEYIYFQTCIRG"
6518. .7825
/gene="VCA0098"
CDS
2400. .2573
/gene="VCA0098"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF6012.1"
/db_xref="GI:9657484"
/translation="MIVTIDMI CURLAPKGIETLLVYRSNPNRBDGIMALPGWVPD
EDLSAOGEPADKDPDSARRICQKITHYPTNFISSDPLVDGNPFGDSGMSISISHA
LINSNVTQIEAEMDSRYVWPALEITLQGVYALPDHVAQIOHAWKILRAAYEYIS
VTVLSIRFETPVADIIETAYAKSGDVARMYTKRLIDTGVIVSNKXLAASKGKGKGP
ATVRLANSNEYIYFQTCIRG"
6518. .7825
/gene="VCA0098"

CDS 6518..7825
 /gene="VCA0098"
 /note="Similar to GB:J05568 SP:P18133 PID:147307 GB:U00096
 PID:1651456; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="nicotinate phosphoribosyltransferase"
 /protein_id="AA096012.1"
 /db_xref="GI:9657484"
 /translation="MNPRLSPHIIIRSLIDDAIKIMQAIIHFPYDVSYRELIYR
 SEDASGLDAIRQELIAGLTFESDADIHVLTQAPLKAFLQSLAFYFVQEOY
 EWGIYQGGKQILIRISWMDITLVEITLWALYSEYRSKAPQSLADPLVLYV
 KLDQKAEIERGINNPSLITEXGRFRRSSQVORVLAELKQSIPOWLTGTSNPR
 BEDLKIGTIAHEWFMGHALVNERDSQVLRULTFDMLAIAPIIDLTIDAFIN
 DPNRLIANYGVRHDSGCPKRMKXNAHQQLGIDPTTCLTFPSGLDLPDQLELC
 EYFAGRKISFEIGFTLNDLANMKNANAVIRPUSIVTKIAECQGRVAKISQPEK
 AMCEDPILANIKRRFNIELVDALIQELRHQKSPHYISAA"
 7856..8896
 /gene="VCA0099"
 7856..8896
 /gene="VCA0099"
 /note="Similar to GB:U00096 PID:1742678 PID:1742686
 PID:1742694 PID:1742718; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="oxidoreductase, Gfo/Idh/Moa family"
 /protein_id="AA096013.1"
 /db_xref="GI:9657485"
 /translation="NPNTPLKGVIGYISAKTFLPDIJLESLTSAISSQOQYV
 EDWFOIAFDSAQRLITESVDVLTITAPNNVHFPLAKLLEHGVIVKEPVTI
 OGRTEILARQGLLSEVFNHRWDGFLVTKLIEGQGLDVGKVRPVTI
 RQRMEOAOEGGILFLDPAHLIDQALVFLPOSLSMDSCDPLDYLQLOVY
 YQHVRLHANYSPBPVRYOVLSGLKYKGLDPEDEMLKGBEPPTHQWQENP
 ROYGLYAEENENITELGGYQHTFYQVVAIRNGASNPSASALQSIQLELALE
 SSAQQLAVTI"
 complement (8933..9304)
 /gene="VCA0100"
 complement (8933..9304)
 /gene="VCA0100"
 /note="similar to GP:7226016; identified by sequence
 22.3%; Score 574; DB 1; Length 10977;
 Best Local Similarity 64.3%; Pred. No. 5,4e-157;
 Matches 1028; Conservative 0; Mismatches 440; Indels 130; Gaps 6;
 gene
 CDS
 Query Match
 Beat Local Similarity 64.3%; Pred. No. 5,4e-157;
 Matches 1028; Conservative 0; Mismatches 440; Indels 130; Gaps 6;
 23 ACTGCATCGTACGCGCGGCGCACTTGCCAGCGCGGCTACGACTTCAAG 82
 6987 ACCTTCAAGTAAATCAAGAGGACCTCAGCCCAAGCTGACGGCTCGTACTTCAAGT 6928
 83 ATTAAGGCCATGACGAGCGTTTCAATAAATGCTGTGCGCCAGCTGACTGAATGCGG 142
 6927 ACGATCGCATGACGAGGTTTCATACAAATCGGTGACGCGCAACTCCAGAAATCTG 6868
 143 AATGCGAGTGGCGTTTGCCCTTGCTTGAGGATCCGATTTCAATTTGCGCATTCGGGT 202
 6867 ATCCGAGTGTGCTGTTTCCCTTGCTTGAGCAATCCCATTTTCACTTGCTTGAGGC 6808
 203 TGAATATGAAATACGTAATGACTGTAAAAAAGTACGATTCMAATGAGTGCATGCTGC 262
 6807 ACAAAATGGAATTAACGACAGATTGCAAGAAAGTGCCTTGAGATGAGGCGATGCTGA 6748
 263 TCTAATAAACAATGTGGCATCCGAAAAGCGCAATGAAGCAACTGATTTATTTCTTGG 322
 6747 GTCAATATATGAATTCGGCATCCGAAAAGCGCAATGTCGAAGATGACTATCTTTGA 6688
 323 CGTACTTCCTTAATAATCGCTATGCTTCATCATCGCGCAATCAATCAATTAAGGCG 382
 6687 CGGATGCGATGAAGAGCCGCTGGCATCTCTTTCACATCGGTACATTAATTAAGGCG 6628
 383 ACCTCAACATCCGATTAACAAGATGAAGCGCTGCGATCATATTGATTTTAAGCATCA 442
 6627 ACCGACATCCGATTAATAAATGCTGTATCGCTTGCGATCATATTGATTTTAAGCATCA 6568

QY 443 AGATCCAAATAAATCCGCGATTAAGAGAGAAATAAGCGATCGCTCATGATGATGCCA 502
 DB 6567 AGATCGAGCAGGCTACCAATGATGCGGAGAAACAGCGGTGATTCATGAT--TGCCG 6510
 QY 503 TCCCTTCGTCGGTTTCATTCATCATTAAGTAAAGCAAGTGTGCTAAC--TTGGG 561
 DB 6509 TCCCTTCGTCACCTTCGGAT-----GTTAGAACATTTTGTGTCACTTTTGCA 6460
 QY 562 CGAACATTAAGTAAACCTTGMAAGTTTGCAACTTTTGCAACAACCTA----- 610
 DB 6459 GGCACGATATGAGCCCTCTAGTTGTGCACTTTGTGCAACAACCAATTAAGAGA 6400
 QY 611 ----- 610
 DB 6399 AGGAATCTCGTTAAAGACAGCCAGCAGCCCTTCAATCCAAACCTGCTGAGTTGC 6340
 QY 611 -----GTCAGTGTATTTGGCCTTATTAATTAATGATAT 647
 DB 6339 AGGTAAGCGGCAATAGATGATCCCATGAGCATAGTACAGATGTATTTGGGTGAG 6280
 QY 648 TGAAGGGTAAAGACGATGATTAACAATTA-----CAGTA 685
 DB 6279 TGAATGTCGCAACACCGTCGAGCTTCAAGTGAAGGATATGATTCAGTCTGCT 6220
 QY 686 CTTCTGTATCTGATGATTTGTTGCAACAAGCTTATTTACATTTGACCATCATGTC 745
 DB 6219 ATAACTGCCCAATTCAGTTATATCACTTTGTTATTTACATTTGACCTTACATTC 6160
 QY 746 ACTTACCTAAATTAAGCCGCTGTTTATTAAGGAGCACTTAATGATGTCATATCGATA 805
 DB 6159 AGTTTACCTAAATA--CCCGCACCTTATGAGGAAACCTTGATGATTTGTCACATTTGATA 6101
 QY 806 TGAATTTGTCGGCTTGGCGCGAATTCATCCAGGTTTACGTTGTAAGCAAGCTTAATC 865
 DB 6100 TGAATCTTTGCGCTTGGCCCGGAGGAGGATTCAGACCTTTGTTGTCACAGCAATC 6041
 QY 866 CAATCGCGCAATGTTGTTAAATGCGCATTTGCTGCGGAGTATGATGACGAAGATA 925
 DB 6040 CAATCGCGCTGATGTCGGCTTTGGCGCATTCGCGTGGTGGGTGATGAGGATC 5981
 QY 926 TGAACCGCTCATGTTGAGAAACCTGTGATGAGATTTTGAAGAGGACGACGATTT 985
 DB 5980 TCAAGCGCAAGGTGCGAACCCTGACCAAAATTTGATTCGCGCGCGGCGCATTC 5921
 QY 986 GTGCGCAAAAAGTCACTATCTTATCTAATTTTATCAGCATCCGCTGTTGATGCAACC 1045
 DB 5920 GCGCGCAAAAATCCATATCTTATCCCACTTATCAGTATCCTTTGAGGACGATATC 5861
 QY 1046 CCAAGCGCATCCGAATGTTGAGTGTGATTTTCCATTAAGCTTTATTAACCCGT 1105
 DB 5860 CTAAAGCGCATCCAGCGGTGAGCATGATATTTCTCATATGCTGCTCAATCATT 5801
 QY 1106 GGAATGCAAAACAATAGAAGATTTTGATGACCCCGAGCGCGCTTAATGTTGATC 1165
 DB 5800 CCAACGTCACGCAATTAAGAGGCTGAATGATTAAGCCGTTAATGTTGTCAC 5741
 QY 1166 TTCTATCTTATCTCAAGAAAGAAATGCGCTGCTTTGATCATGTCGCGCAATTAAC 1225
 DB 5740 TAGAGACGATTTTGAAGGTAACAATGCTGCGCTTATCATGTCGCGCAATTAAC 5681
 QY 1226 ATGCGTGGCAAAATTAACCGCTGCGTGAATACATCCGTTGACTATTTTCACTAG 1285
 DB 5680 ATGCAATGAAATAAATCGCGCGCGCTGAGTACCTTGTAGTGTGCTTCTCATCTG 5621
 QY 1286 AAAAAGTTTTTAAAGTGGGCAATTTATGATGCCATACCAATTTGCGCTGAAAGTTA 1345
 DB 5620 AGCTGAGTTTTTGTGGCGGACATATTGAGGCTTACCCAAATTTGAGTGTGAGTGA 5561
 QY 1346 ATGCAATGACATTAAGCGCTTGAATCAATTAACGGGTTGATGTCATGACATCAATAA 1405
 DB 5560 ATGCAATGACGATTAAGCGCTGATGATTCGCGGATTAATGTCATGACATTAATAA 5501

QY 1406 TGCGCATCTGTGAAGGCAAGCAACGACCGTTATGCTTGCAGTC 1465
DB 5500 TTCCCGCTCTACGAAGGTAAAGCGCAACCTGCGAGCGTTATGCTTACGAGTA 5441
QY 1466 ATGAAGTCACTATTTTCAACCTGTTTACGAGTTAA 1503
DB 5440 ACGAAGTCACTATTTTCAAGCTTGCTTGCAGTTAA 5403

RESULT 2
AE016805 306147 bp DNA linear BCT 29-SEP-2003
LOCUS Vibrio vulnificus CMCP6 chromosome I section 9 of 11 of the
DEFINITION complete sequence.
ACCESSION AE016805 AE016795
VERSION AE016805.1 GI:27361837
KEYWORDS
SOURCE Vibrio vulnificus CMCP6
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 306147)
AUTHORS Kim,Y.R., Lee,S.E., Kim,C.M., Kim,S.Y., Shin,E.K., Shin,D.H.,
Chung,S.S., Choy,H.E., Progulake-Fox,A., Hillman,D., Handfield,M.
and Rhee,J.H.
TITLE Characterization and Pathogenic Significance of Vibrio vulnificus
Antigens Preferentially Expressed in Septicemic Patients
JOURNAL Infect. Immun. 71 (10), 5461-5471 (2003)
PUBMED 14500463
REFERENCE 2 (bases 1 to 306147)
AUTHORS Jeong,H., Moon,Y.H. and Kim,J.J.
TITLE Direct Submision
JOURNAL Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
Yuseong-gu, Daejeon 305-811, South Korea
REFERENCE 3 (bases 1 to 306147)
AUTHORS Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.
TITLE Direct Submision
JOURNAL Submitted (13-DEC-2002) Department of Microbiology, Genome Research
Center for Enteropathogenic Bacteria, Chonnam National University
Medical School, Hak-1-Dong, Dong-gu, Kwang-Ju 501-746, South Korea
FEATURES
source location/Qualifiers
1. 306147
/organism="Vibrio vulnificus CMCP6"
/mol_type="genomic DNA"
/strain="CMCP6"
/db_xref="taxon:216895"
/chromosome="I"
92..1075
/locus_tag="VV12370"
92..1075
/locus_tag="VV12370"
/note="COG0016"
/codon_start=1
/transl_table=11
/product="Phenylalanyl-tRNA synthetase, alpha chain"
/protein_id="AA010744.1"
/db_xref="GI:27361838"
/translation="MHIQDEITANANAAIDAQSVVALDEVRYVGLGKGETLTQLOG
LGLPPEERERAGOEINRAKEVVAHALAKDALRALEAKLASETIDVLTPEKRIE
NGKLPHTVTERIEQFPEGLFENVSPEIEADPHNDALNIADHPARDHDTFFR
NFDLMKRTISQVQIRTMNGKSPFRFLAAGRVYNDYDQCHTFMFGVEMLVDEVY
NFAOLKGIHDFLCNPFEEVEVERFRPSYFPETTESAEDVKGKNGKMLEVTLGGGVNH
PVLRSVGIDPEKSGFAPGCMGVERLTMLRGVNDLRAFPENDLRFLEKQPK"
1094..3481
/locus_tag="VV12371"
1094..3481
/locus_tag="VV12371"
/note="COG0072"
/codon_start=1
/transl_table=11
/product="Phenylalanyl-tRNA synthetase, beta chain"
/protein_id="AA010745.1"
/db_xref="GI:27361839"

gene
CDS
/translation="MPSSESNLRKEMNPVATTDDELTHQITAGLEVDVLTPLPAQTENG
VKTGVVEGQHPDADKLVTYKVDVEEELLDIVCGAANCROGLKAAVAVAVAGD
FKTKKAKLRQPSGHWGLCSFTGLGIDVEDSDIMELAPIDPGLDFEDFDLADLVYDV
DLTSNADQFISIGMARVEGVNADVPSPVAPVSDIDVVAIEVKAPAAQPRQIG
RVKVNVOQKTPILMMQEKLRGCGISIDIPVDITVFVLLGQGPMAHDLAKIOGI
VVRALGQEKITLLDGSBAELNADTLVADHKKALAIAGIPGSESGVSEKDVYLE
CAFPADHIRGRARSYGTLTDSMREPRGVVALQVMSAMERLALLVEICGGEVAPV
AYSEBALPREKNVMAKRTKLDNLGHHIADSDVBEILERLMTVETTABGVVAPPT
WRFDAIEQDLVEGRITGYDITPQNPAAALKNHDOENIPLKRVADLVNKGTH
EALTYSFVEPEQKLVPGVDALLFENPISAEKMSARLQLQGLNVTVMHNRQQR
VRLFEYGLRFIPCDITENGRQEPMLAGVIAIGRSEHNIDITNVDFDLKGDVAIL
LELSANDKAVSFVAAGPALHPQOSAIIVDGEKIGISGVTEBERKSGVSEKDVYLE
EIEWASINKRVIPEAVALSKPANREDIIVVDEAVASGDIVNACIEVGEFFLKATL
FDVYVGKVGVEGKKSLAIALTLQSNRTIEDDADIAGVDAIYTHSEKSGALRD"
complement (3584..4897)
/locus_tag="VV12372"
complement (3584..4897)
/locus_tag="VV12372"
/note="COG1488"
/codon_start=1
/transl_table=11
/product="Nucleotide phosphoribosyltransferase"
/protein_id="AA010746.1"
/db_xref="GI:27361840"
/translation="MCAPLFOPAIIQSVLDLDPYKINMQAAVRYEPQOTQVRYELLIV
SDNLSDLVEEVEEINRLAEFDQAQALAEKAPVLAELSYLETFRFPQOV
SVGIFRTAQQDCOLRVINGIWHETILYELVMSIISLRNRRVMAOIPOSOULKYLE
KIDLPFDSLEKRTNTPRPSMGTCRRSPFAOKMVLVLAARVLELITGNSHYLA
QEPNLPITGVVAHERTMQAQALVALIHSRVALDKRLBAPNSLGALNTDITGIDTFL
SDPLDKATRVAGVRHDSGPVWGDKITLAHESLSDIDPTTKLITFDGIDPRARDI
CEYFAPRAQISFGITGLANDMGNWTNDGTRQPTISWVNAECNCSVAAKISDEPE
KAMCEDIFIFLMNKKORFGLGELVDLKAILETKMKRQKRKIOSVA"
5176..5889
/locus_tag="VV12373"
5176..5889
/locus_tag="VV12373"
/note="COG1051"
/codon_start=1
/transl_table=11
/product="ADP-ribose pyrophosphatase"
/protein_id="AA010747.1"
/db_xref="GI:27361841"
/translation="MIVTIDMTCTKIGKGLSVTLKRNPRPHQMSIPGCFVFE
HDSQGGQPADADPSARRRICRQKITHTYPTITSPKVDGNPKRDPKMSVTIAYIA
LNLQNTIEQIQCNGCLDQQLAMPALSVQGVALLAFDADLIRLMTLRAIETIS
VILFALKEFLVSDIITHAYAKGVVNMRTIKRRLIDTGVIVSANKIASTNKGKGP
ACVYSIADKHVTFQTCR"
5902..6546
/locus_tag="VV12374"
5902..6546
/locus_tag="VV12374"
/note="similar to nicotinamide; COG1335"
/codon_start=1
/transl_table=11
/product="Amidase"
/protein_id="AA010748.1"
/db_xref="GI:27361842"
/translation="MVDIKKRTASLDVDPQGFSELCENELPVSGALEIVKAKNH
TKAKLUVSRDHPGAAAEATPAAKMEVGLPEVDIKWNPACIVGTGVELDPELQ
ALREYDFQNMKGIDPDPAHVGAFYHQADLTSTGIEFLRARIDVAVGLGALDPCV
KKSIEQLVDAFEVILNLSAPVFPENNQVAVSHLTALGVICIDVNOQLHE"
6962..7249
/locus_tag="VV12375"
6962..7249
/locus_tag="VV12375"
/note="COG0776"
/codon_start=1
/transl_table=11
/product="Integration host factor, alpha subunit"
/protein_id="AA010749.1"
/db_xref="GI:27361843"
/translation="VALTKADLAENIFELGSGSKDAKTVVFPSEIRKALENEQV
KISGFNPDLRDKNERPGRNPKTGEDIPITARRVVTTPRGQGLKARVENLK"

gene
complement (7347, .8123)
/locus_tag="VV12376"
complement (7347, .8123)
CDS
/locus_tag="VV12376"
/note="COG2357"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AAO10750.1"
/db_xref="GI:27361844"
/translation="MSLFRTALMLVLSRAPAPAIPTAPSNOKORETSOREVCS
KLFKSLGIVGISTDLOPOTPDYDPRVLSKAOQAFELITVKSFTALLTDOPAF
AGVSGQAKOKIALIELNNOPERTIDLRATTVADVASLVSATBALRETRITVKN
RKSFGSGYRDNLVLVQDPTNLVAEVQLHKAIDVSGSPEDHLYOQIQWEMROAA
SETRLNDVELATITRLSRSHKDIYQAMOPYLTTHLEAA"
8413, .3066
gene
/locus_tag="VV12377"
8413, .3066
CDS
/locus_tag="VV12377"
/note="COG0500"
/codon_start=1
/transl_table=11
/product="Thiopurine methyltransferase"
/protein_id="AAO10751.1"
/db_xref="GI:27361845"
/translation="NRDAPFMNKKASNOIGFHEEDVNPILTRMSALAPREETVY
PLCGTEIDAMLATKHDEEGAEISLIRSFPAHFTPTVPTISGHEHYOPDEIS
IYAGDFTAPLSKADLIYDRAIALPBEKVEYVQRIKGLNKGGRLLVLSYDPO
EMAGPFSVTOGEIHLFAGMHTLXYODIADHEHPKIAKQSLRFSBEVYIENDY"
complement (9119, .10294)
gene
/locus_tag="VV12378"
complement (9119, .10294)
CDS
/locus_tag="VV12378"
/note="COG0027"
/codon_start=1
/transl_table=11
/product="Phosphoribosylglycinamide formyltransferase 2"
/protein_id="AAO10752.1"
/db_xref="GI:27361846"
/translation="MGTATRSATRVLLSGELKEVAIECORLGEVIAACRYPD
APAMVARSYFDMLESELKVIAPAPVPELTAITDKLVEIEEGALNVVS
AKATLTNRREGIRLAEELGTTSPFRADNYQOFVEAVASIPVAVSSSG
KGQSVTKSPADIEKAMOYAOBGRGARVIVEGIDVIDYITLLVCAVADGVCAP
IGHROEDDYRESNQPMQSENAIKAAVTAEOVNAVGIGIVELFVGDVIRN
EVSRRPDTGLVTLISQMSSEPALHRAFTGAPVVKIYQSPASAVILIGQSENR
FDGMSDALBPOTOLRFKXPDINGRRRLGVTLRRSETEKAVAAIESAKIKITV"
complement (10426, .11334)
gene
/locus_tag="VV12379"
complement (10426, .11334)
CDS
/locus_tag="VV12379"
/note="COG0295"
/codon_start=1
/transl_table=11
/product="Cytidine deaminase"
/protein_id="AAO10753.1"
/db_xref="GI:27361847"
/translation="MOLQFTKRSFENALASAPESLRHLAPITLAEPDASLSTVO
PEILKOTGMNDNLVALIPRAAYSAPISEFYGAIVLGLGALYFGAMFEDVY
ULGOTVHAOSASISHAMKGGQSLDITINSPCGHQFNNELSSAKELKIQPEER
EKKHIDVLPDSRGPDLIGEGALMSOVHGFATDDDLMDRAVAMNRSHAPYTHNL
SGVALQTSRGRTICAYANNAFNLSLEPIQVALIQLILAGERFENIQSALVESHKG
KISHLACTQSTIEALNPDIPVSYLSL"
complement (11675, .11812)
gene
/locus_tag="VV12380"
complement (11675, .11812)
CDS
complement (11675, .11812)
Query Match 18.1%; Score 467; DB 1; Length 306147;
Best Local Similarity 59.1%; Pred. No. 3.3e-125;
Matches 864; Conservative 0; Mismatches 590; Indels 23; Gaps 4;
QY 5 TGCACCTTATCAGCAATACCTGCACTGCGGCGGCACTTGCGCCAGTGGCGG 64
Db 4407 TCCAGTTTCTCTTAACTCTTGCACTGAGACTGGGGTATTTGTCGCCAATTAACGA 4466

QY 65 CGGTAAGTACTCTCAAGATTAAGGCAATGACTAGCGTTTCATATAAATGCTGTCCG 124
Db 4467 CGATTGGGAGGCTGTGAGATTAATCACTATCACTAAGGTTTCAATAGAAATGCTGTGCG 4526
QY 125 CAGTACCTTGAATGGCGATAGCGAGCTGGCGTTTGCCCTCTTGCTTAGG--ATCCCG 181
Db 4527 CAATGCGATTATGAGTCTGAGCTGAGCTGCGAGCTGCCCTGTCGCGTGCAGAAATACCG 4586
QY 182 ATTTCAATTTGCGGATCGGGTTGAAATGGAATAGGATGATGATGATAAAGATACGA 241
Db 4587 ACAGAGCTTGTGCTGAGGCTGGAAGCGGAAGTTTCTTAATAGAGAGAACTCTGCC 4646
QY 242 TTCAATGAGTGCATGCTGCTCTAATAATAACATGTCCGATCCGAAAGCGCAATGAA 301
Db 4647 GTTAAATAGGCGCTTCTCTCGCAAGTAAAGCAATGGGCAATCAAGCAAGCTCT 4706
QY 302 GCCAATGATGATTTCTTGCGGTAATCTTCTTAATTAATGCTAATGCTTATGATCATG 361
Db 4707 GCGAGCGGTTGATTTCTTCAAGAACCTTCTCACTAATCAGAACATGTTATCATCAGAG 4766
QY 362 CGCAATCAATTCATAGCGCACTCAATCCGATACAAAGCAATGAAAGGCTGCATC 421
Db 4767 CGAAGATCAGTTGATGAGCGCACTTGCTTGAAGATAAATGATGATACGCGCTTGATC 4826
QY 422 ATATTGATTTTAAGGATCAATGATCAATTAATCTGCGATTAAGAGAGAAATAGG 481
Db 4827 ATATTGATTTTAACATCTAATTAATGAAGAACCTTGAAATATGCTGTTGAAATAGT 4886
QY 482 CGATCGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 541
Db 4887 GCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4929
QY 542 AACGTTGCTTAATCTTTGGCGCAATTAAGTACCTTGTAACTTTGTCATCTTTGTCG 601
Db 4930 ATTTGCTGCTTATGATGCTGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTT 4988
QY 602 ACAACCTAGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 659
Db 4989 ACATACCTCAATGCTGCTGCGAGGATGATGATGATGATGATGATGATGATGATGATG 5048
QY 660 ACGTATGATCAACAAATTAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
Db 5049 GATGAGCGGCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5108
QY 720 TTATTTACATTTGACCATCATGATGATGATGATGATGATGATGATGATGATGATG 779
Db 5109 CTTTCCATTTGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5168
QY 780 AGCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 839
Db 5169 TTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5228
QY 840 GCTTTTACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 899
Db 5229 AGTCTGCTGATTAAGGCAACATCTGGAACACACAGATGAGATGAGTGGCTTATTC 5288
QY 900 TGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959
Db 5289 GGGAGGATTTGTGTTGAACAGATTTATCCAAAGGGAGGACCAACCGCTGACGCTGA 5348
QY 960 TTTTGTATGACGAGACAGAGTATTTGTCGAGCAAAAGTCCATCTTATCTTAATTTAT 1019
Db 5349 TTTTGTATGACGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5408
QY 1020 CAGCGATCCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1079
Db 5409 CAGTGAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5468
QY 1080 TTCCCATTTACGCTTTATTAACCCGTGGAATGTCAAACAAATGAAGATTTGATGATCGA 1139
Db 5469 CGCCCATTAATGACCTCTTAATCAACCAATATGAAACAAATTCATTAATGATGATGATG 5528

| | | | |
|----|------|---|-------|
| QY | 1140 | CCCCAGCGCGCTAATGTGGTTGATCTTCATACCTTTACTCAAGAAGAAATGCCGTGGC | 1139 |
| Db | 5529 | TGATCAACAATTTGGCTTGCTTCCGCTTGAGTCGCGTGTCTGCAAGGGGAAGTTGCGTTGGC | 5588 |
| QY | 1200 | TTTGTATCATGTCCGCCCAATTCAGACATGCCGTGGCAAAAATTACGCGCTGGCGTTGAATA | 12585 |
| Db | 5589 | CTTTGACCATGCCGATTTAAATTCGGCTCGACATGACCAAGTTGCTGGCGCGGATCGAATA | 5648 |
| QY | 1260 | CACATCCGTGTCACATTTTTCATTAGAAAAAGAGTTTATGCGCGGATATTAATGATGC | 1319 |
| Db | 5649 | CACATCGGTGGTCTTGTTGCACTCGAAAAAGATTTCTCGTTTCTGACATCATTTCAAGC | 5708 |
| QY | 1320 | CTAGCCCAAAATTTGGCGCTGGAAGTTAAATCCGATATGACATTAACGCCGCTTGATCAATAC | 1379 |
| Db | 5709 | CTAGCCCAAGTTGGTGTGAGATGTAAACGCAATGACATCAAGCACTTTGATCGAATC | 5768 |
| QY | 1380 | CGGGGTGATTCGTCACTATCAATTAATAATGACCGCATCTTGTAAAGGCAAAAGACCAATC | 1439 |
| Db | 5769 | CGGTGTATTCGGAACGCGAATTAATAATTTGCTCGACCAACAAAGGGAAGGACCAATC | 5828 |
| QY | 1440 | AGCCACCGTTATTCGCTTTGCCAGTCATGAAGTGCATATTTTCAAACTGTTAAG | 1496 |
| Db | 5829 | CGCTAGATCTATAGCTTAAGCGGATTAACACGTAACTATTTTCAAACTTGCTTGGC | 5885 |

gene
CDS

1151..2908
/gene="VV1707"
1151..2908
/gene="VV1707"
/codon_start=1
/transl_table=1
/product="methyl-accepting chemotaxis protein"
/protein_id="BAC94471.1"
/db_xref="01:37198637.1"
/translation="MSGSPVQKRFPPFSIRAKLLINVTLLGVAAYAIYENNSAGH
LSSEHASTENLSIDLLMRHEKDYIARSRKVTSPDATYCLQCGIDHLSVDYL
ASESIALDORCOQIVLTLSYQVQHOLAQLGINDLQSELPFARMELKTNVLSRV
GELRAQFOLDHDLITVTRPTQHKVALLGLSQPKFSAADLSIPATDYNTLSRV
SAITTELTITNEGRGRLRSRVHTBEATIDLTCTTQGVQFQRAADIRSLQAMGVYI
VLLSLSLTYIGRSILSRKAINLMDIDANGSDLTFRNNAKSDSLAQLSRSDFL
TSHLADNIKQLASVDYLVGSSCSSEQAPARSQWNAQVSEADKALQTHAGKRD
TLAISIESQAOQLALAQSRREINAVVSAIQGLAQTNLALNNALEAAGSRGSE
AVLADEVQVQSRLENTDSTLQIESVCAITKGIGIQTNRKADSVQATNTVDTRHVE
AIEGISTQITTEMFMTQIATASEECQSMVAEIDRNTQIQAIGDYTEIVSGVRS
EYQSDVSKTEIEVAQPKY"
complement(3018..4610)
/gene="VV1708"
complement(3018..4610)

gene
CDS

| | | | | | |
|------------|---|-------------|--------------|-----------------|-----------------|
| RESULT 3 | AP005337 | 241900 bp | DNA | linear | BCT 04-DEC-2003 |
| LOCUS | Vibrio vulnificus Y016 | DNM | chromosome I | complete genome | section |
| DEFINITION | 8/14. | | | | |
| ACCESSION | AP005337 | BA000037 | | | |
| VERSION | AP005337.1 | GI:37198635 | | | |
| KEYWORDS | Vibrio vulnificus Y016 | | | | |
| SOURCE | Vibrio vulnificus Y016 | | | | |
| ORGANISM | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio. | | | | |
| REFERENCE | 1 | | | | |
| AUTHORS | Chen,C.Y., Wu,K.M., Chang,Y.C., Chang,C.H., Tsai,H.C., Liao,T.L., Liu,Y.M., Chen,H.J., Shen,A.B., Li,J.C., Su,T.L., Shao,C.P., Lee,C.T., Hor,L.I. and Tsai,S.F. | | | | |
| TITLE | Comparative Genome Analysis of Vibrio vulnificus, a Marine Pathogen | | | | |
| JOURNAL | Genome Res. 13, 2577-2587 (2003) | | | | |
| REFERENCE | 2 (bases 1 to 241900) | | | | |
| AUTHORS | Chen,C.Y., Wu,K.M. and Tsai,S.F. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (29-MAY-2002) Shih Feng Tsai, National Health Research | | | | |

es. 13, 2577-2587 (2003)
s 1 to 241900)
, Wu, K.M. and Tsai, S.F.
mission
d (23-MAY-2002) Shih Feng Tsai, National Health Research
Division of Molecular and Genomic Medicine; 128,
Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China
gertsai@nhri.org.tw, Tel:886-2-8146-1041,
2-2-2789-0484)
ence was determined by the Sequencing Core of the National
University Genome Research Center (YMG;
genome.yu.edu.tw).
Location/Qualifiers
1. 241900
/organism="Vibrio vulnificus Y7016"
/mol_type="genomic DNA"
/strain="Y7016"
/db_xref="taxon:196600"
/chromosome="I"
197. .1024
/gene="VV1706"
197. .1024
/gene="VV1706"
/codon_start=1
/transl_table=11
/product="formate dehydrogenase accessory protein"
/protein_id="BAC94470.1"
/db_xref="GI:37198636"
/translation="WVRKNTLITSENPLOTLEVERFDEYGEKLTIKQACERPLTILN
KKEVITLITLGRPEALVIGYIKKDSFLSEYDSESVITIMETNSAAVYETENHEIT
ALKKRVITSGGGCGTGMGNWKKQDEHYKVPQVTLKQSEIYKTLBALRHVDYTKKAG

YAMGAVCVLGGSDSEVSEVDEGVGHNAVDLTLAGEMTNLQESGADPIPTTGGTLTSEMTVY
 VAMGAPVLVLSGSEVTVGMGLDLARQCITTIARAKLRQRVFGSGENIIIFDVAGQGS"
 1151..2908 /gene="VV1707"
 1151..2908 /gene="VV1707"
 /codon start=1
 /transl table=11
 /product="methyl-accepting chemotaxis protein"
 /protein_id="BAC94471.1"
 /db_xref="GI:37198637"
 /translation="MSGSVQGRPFPSFIRAKLLINVTLLIGVAAYILENNASAH
 LSELSKHAETSNSSSITDIIRHEDVDIARSEKTLTSPDAYSOQSDIHLSDVL
 ASSEIADIDROCOIYTLISQYQYQFQGLAQLANIDALQSELAASMLKTNVYLSAND
 GELTKPQFLQDLHDITPEVIRPQEHQVALLDGLSQPAKPEALDLPFTDYSNTLSRVY
 SMTITGELTANGSGALASNNHKTGEALTDITTKQCEVQRTAIDRLHGMGVYI
 VYLLSLLYTIRSLISRKAINLMDIDANSQGLTYRNKAGSEMLQSSFDLF
 VLSLHDNIKQLSWDVDESSCSQAQSNQNRVNSEDKALQTLHAGKID
 TLAIIEESQALQLEAQRSSINVAIIQGLAEQTNLALNALAEAAQSGSGSF
 AVAGDEVRLSTETMDTQTLISITVQALTKGLAKNKAQSDVQARTINVDPRVIE
 ALEAGSTQLETFEEDMTQTLTASQESQWSAETDRNITQIAGLQGYEIVSGVCS
 EGVDSVSHLEKTVQAFK"
 complement(3018..4610)
 /gene="VV1708"
 complement(3018..4610)
 /gene="VV1708"
 /note="containing duplicated Arpase domains"
 /codon start=1
 /transl table=11
 /product="ABC-type transport system, Arpase component"
 /protein_id="BAC94472.1"
 /db_xref="GI:37198638"
 /translation="MISTANIQQGAKPLFENISVKGEGNRYGLIGANGCGKSTFM
 KILSELEBTSQNVSDPYERVAKINQDQFAEEFVIDVTIVNGHEMLMVQGRRI
 YLPLSEMSDEGRVADIEVEFPMQDYMALAKGELTAVGIMEQHYGMSVAGM
 KIRVALLQELPADPIMLDEPTNINMDPTIRLBEETLANQNTMIIISDRPLNSV
 CTHMDIDTGLFPRVPGVDEWTALQSARRELLSNARKKQIALQTFVARFSNA
 SKRQKTSRAKQIDIKLDEYKASQKNPFIRPQSKELPRNALITYENISQGEEDLF
 SDFNAIFVEGEVVAIIIGENGQKTLTLNLAGLVEPTGSPKASNNATGYTAQDPAH
 EAEDNLINEMWQVQRESDQVKGFLGRMLFGQDIIKQSVKVLISGGQGMLLGK
 LMAHPRMNLSEYLNKSMESIESLNALBOYKGLFVSHDRVVDLSATRIIEIR
 DGIITDFKQTYSEYLNKSGES"
 4834..7179 /gene="VV1709"
 4834..7179 /gene="VV1709"
 /note="identified by Genemark and Glimmer2"
 /codon start=1
 /transl table=11
 /product="hypothetical protein"
 /protein_id="BAC94473.1"
 /db_xref="GI:37198639"
 /translation="WQKLRILVCFVLVAPATFQMOYAAQTPAVQSRQALISESL
 SALPPLFEMGPPEHKKVNDLAVVELQTEGGAFAAQAQYKAPAPQWLIVENYD
 LTNISGSSKAKDLQITDITASBGLQFGPBYGTQPROBELAKNLEIYVHQLRSEF
 RLAKINTFSPPVPLWTGSAVAVYLLVMMNASKRIITLFEESQLENTAPLWPK
 FWTYSRAKAKLAAIIVSRILASLISLOHLYLFTWMLIGSIIISLIEFA
 VYNSGSSSDVALVATIRVYMSLVAGVILQISVLTIGKTIYWIISALYEMFA
 LTVSYVLRMKKVEVDRIETRPYVQMAIKRQSPFLINTPATAIAWMLSHHKK
 HIIILNTSNYMSQLALFPIEAKQTSNRDGNLSVIRIDPAFOYLPVQPODE
 WIDVASDEQNSLRVLTDSPAICVLSGEGVQTSVLKIKLKVKNAPLVLPQOBE
 GPPELLKQALQALGAESTVQYIQLFLKRSRPVYIADNQRVLQVPMWGLAGLWK
 LTNLRLSKQNRVMAIEKASVQIFDARSGEELLIDWYSFLRNKKQSLIADLSQK
 NQDIEMLPSFDGLVYDEKQELTSEEEKQAFYIILWYSGNFTVALPRLSLR
 NKEDSSVYVRLFHAPEKSELQMPKMLAVIRISYQLEVASBELSECTQLSIAEVI
 GTLRPFESGCFLANNEDKRVRDHWRIYITVNLDRGLLVK"
 7181..8062 /gene="VV1710"
 7181..8062 /gene="VV1710"
 /codon start=1
 /transl table=11
 /product="small-conductance mechanosensitive channel"

| | | | | | |
|------------|--|--|--------|----------|---|
| QY | 900 | TGCGGGATATGTATATGACAGATATGACCGCTCATGTGTGAGAACTGTGTATGAGGA | 959 | TITLE | Direct Submission |
| Db | 201543 | GGGAGGATTTGTGTTTGAACACATTTATCCCAAGGAGGAGACACCGCGTAGCGCTGA | 201484 | JOURNAL | Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp, Fax:81-6-6879-2047) |
| QY | 960 | TTTTGATGACGACGACGATTTTGTGTGGCAAAAAGTCGCAATCTTATCTTATTTAT | 1019 | COMMENT | This clone was isolated from a patient presenting with acute gastroenteritis |
| Db | 201483 | TTTTGATTCGCGCGACGCTGTCTGTGTGGCAAAAATTCATCTATCTCATACAT | 201424 | FEATURES | Location/Qualifiers |
| QY | 1020 | CAGCGATCGCTGTGTGTGACCAACCCCAACCGCATCCGAATGTGTGATGTCAAT | 1079 | source | 1. 303450 |
| Db | 201423 | CAGTGAAGCCATGTGTGTATGTATGTAACCAAAACCGCATCTCTATGTGTGAGCGTACAT | 201364 | | /organism="Vibrio parahaemolyticus" |
| QY | 1080 | TTCCCATAGCGCTTTTAAACCGGTGATGTCAACAAATAGAAATTTGTATGCA | 1139 | | /mol_type="genomic DNA" |
| Db | 201363 | CGCCCACTATGCTCTGTATATCAACCATATGCAAAATTCAAAATCTGTGTGG | 201304 | | /strain="O3:K6" |
| QY | 1140 | CCCCGAGCGCTAATTTGTTTATCTTCACTTACTCAAGAAAGAAATGCCGTGGC | 1199 | | /sub_strain="RIMD 2210633" |
| Db | 201303 | TGATCAGCAATTTGCTTGTTCGCGCTTGTATTCGTGTGCAAGGAGGAGTGGCGCTGGC | 201244 | | /db_xref="taxon:670" |
| QY | 1200 | TTTTGATCATGTGCGCAATTCAGCATGTGTGGCAAAAATTAACGCGCTGGCTGAATA | 1259 | gene | /chromosome="2" |
| Db | 201243 | CTTTGACATGCGCATTTAATTTGCTGTGATGACCAATTTGCGCGCATGCAATA | 201184 | CDS | 81. 1574 |
| QY | 1260 | CACATCGGTGATCTATTTTCAATTAAGAAAGATTTTATGCGCGATATTTATGATGC | 1319 | | /gene="VPA0300" |
| Db | 201183 | CACATCGGTGATCTGTGTTTCACTGCAAAAAGAAATTTGCTGTGCAATATTCAGC | 201124 | | /note="similar to EMB:CA6187.1 (AL591787) percent identity 68 in 487 aa" |
| QY | 1320 | CTACGCCAATTTGCGCTGCAAGTTAATGCGATGACCATTAACGCGCTGTGATCATAC | 1379 | | /codon_start=1 |
| Db | 201123 | CTACGCCAATTTGCGCTGATGTGATGACCCGATGACATCAACGCGCTTATGACAC | 201064 | | /evidence=not_experimental |
| QY | 1380 | CGGGGTGATGTGATGACCAATTAATGCGCGCATCTGTGAAGCAAGACCAACC | 1439 | | /transl_table=1 |
| Db | 201063 | TGCTGTATGTGATGACCGCAATTAATGCTGTGACCAACAAAGGAGAGGACCAACC | 201004 | gene | /product="hypothetical signal peptide protein" |
| QY | 1440 | AGCCACCGTTATCTGTTGCCAGTCAAGAGTCACTTATTTCAACTCTTTACG | 1496 | | /protein_id="BAC1643.1" |
| Db | 201003 | CGCTCAGGTCTATAGCTTAGCGATTAACACGTAACCTTCAACTGTGTGG | 200947 | CDS | /db_xref="GI:28808466" |
| RESULT 4 | AP005085/c | 303450 bp | DNA | linear | BCT 05-MAR-2003 |
| LOCUS | Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6. | | | | |
| DEFINITION | AP005085 BA000032 | | | | |
| ACCESSION | AP005085.1 | GI:28808465 | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |
| TITLE | | | | | |
| JOURNAL | | | | | |
| MEDLINE | | | | | |
| PUBMED | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |
| TITLE | | | | | |
| JOURNAL | | | | | |
| MEDLINE | | | | | |
| PUBMED | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |

```
identity 54 in 243 aa
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative regulatory protein"
/protein_id="BAC61646.1"
/db_xref="GI:28808469"
/translation="MKPRQNELQVYDRKAVQVTELSDTIGSGVTIRQDNLTEQ
GYLRVHGATLQSDSDIDTRLEVRFDIKQTLANKADLVAPNETVILEGSSANALLA
RTIARGDVTTIITPSAYIAHLIRNTSANIILGGVYQHGSLVGLPKLCIENIHS
TAFIIGDHQDTGTSTSRDMKRAIDAEAILAKKRNIVLTDSKFGQIYPSISIKTNE
ISVLTDKRAFPSDLEOLKLGVEVVLG"
/complement (3733, .4089)
/gene="VPA0304"
/complement (3733, .4089)
/gene="VPA0304"
/translation="similar to GB:AAF96321.1 (AE004376) percent
identity 76 in 117 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC61647.1"
/db_xref="GI:28808470"
/translation="MHQEKLVYFETGPNIGATKAFPEHYFGMFPVYGPDYAFSG
OGLDGFYSAECVQSTNSGALLVYSSDIHATLEKAKKFGGQIIRPIPEPGCRFH
FTEPGNPAVMSSEH"
/4319..5860
/gene="VPA0305"
/4319..5860
/gene="VPA0305"
/translation="similar to REF:NP_231225.1 (NC_002505) percent
identity 83 in 510 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="catalase"
/protein_id="BAC61648.1"
/db_xref="GI:28808471"
/translation="VSTHWRATQMSKSFLLITVGLASTSLQNTLRDNGAFVGN
ONSITAGENGSVLTQDVLIOKLPFAERIPEVVARAGTGAGSFPVAGSPDLTV
SAPTEKSKVTVFVRFSTVTHSKSPETLDPGPRATKFTTEGMDLVGNLPIVF
IRDSIKFPDMVHSLKPSPTVIVQDPNRFDFSHSPKTHMLTVNLTGPASRTM
DGFVHAQWINOQDVNVKFOVMSQGIKSLRPNTVMQOGDFMLTLNLTVAALG
KGNVPMKLVYVYSPALSKLDYVGLDATTVMNLVDPKVTMLRLBENFLEIE
OSAFAPSNLIGIESEEDRLQGLFAYADTQLYRLGANLFQLPVNRPLSVNNHON
GLSNNAOLSNQDVNVEPSRKXNLAEENQFAVEKLVGTVOOKAISKPRPYOAGVY
RSMNQDRSDLIANLADINKYIDKDIKATVVSIFYRADKETSRLAARDTNLSQVK
NKAMM"
/5962..6429
/gene="VPA0306"
/5962..6429
/gene="VPA0306"
/translation="similar to GB:AAF94738.1 (AE004235) percent
identity 63 in 152 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="ank protein"
/protein_id="BAC61649.1"
/db_xref="GI:28808472"
/translation="MKNYFLLSVLFSGLSFSPLASODKLETOYOSIVLFPDARI
GNBYVETPVQGFIDORNNESYTAALVAYQNKOTVRLIDSGNACLDQKRGTT
ALMGALIKREISIALDLYQACSPDLRKAGLDLKTAEIYQSEVLSKQNMK"
/complement (6548, .7201)
/gene="VPA0307"
/complement (6548, .7201)
/gene="VPA0307"
/translation="similar to GB:AAF42465.1 (AE002563) percent
identity 47 in 200 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
```

```

gene
CDS
/translation="putative pyrazinamidase/nicotinamidase"
/protein_id="BAC61650.1"
/db_xref="GI:28808473"
/translation="MTVRVYQTTAVIDVDPKFSFELCPDELVPVAGALDIVBELKN
HAKGRKLIVSRDLHPKAAWDATPANHCHLPTGTVGELDGL
PPLVDYDFQVNMKMDPAHPYGIFFHDVADTTPGQVNSBELKCKIDTIVGGLADRC
VKSIVQALDLGSKVITVNAATRAVLPDVTDSVIAMKKGALFVNADITVERPA"
/complement (7234, .7959)
/gene="VPA0308"
/complement (7234, .7959)
/translation="similar to GB:AAF96011.1 (AE004352) percent
identity 60 in 241 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC61651.1"
/db_xref="GI:28808474"
/translation="MTVYDDIIPRUSGADKGLBYLTKRSNPRPYHWALPGFP
VPRKDTSEGRPADENFPAARIRIKRKHITPRFSAFIDGKRPEDSLIT
HYALVDRNVEQIINAGVECOLKMPPLQIINGETTLAFDQKTEKAMQKIRASIE
YISVLLFADKEFLVADIIISAYQEPGIDISMTIKRRLIDSGVLKFTNRVASTNKKG
GKPAVYTLTSDSEVTFQNCIRG"

Query Match      15.6%; Score 403; DB 1; Length 303450;
Best Local Similarity 56.1%; Pred. No.2.4e-106;
Matches 860; Conservative 0; Mismatches 640; Indels 33; Gaps 4;

QY      6  GCACTTATGACCAATACCTGATCGTAACCTGCGCGGCACTTGCGCCATGCGCGC 65
DB      8748  GAACCTTGCTTCATACGATTGGAACCTGCTGCAACGGGATATCGAACAGCTTGAC 8689

QY      66  GGCTACGATCTTACAGATTAAAGCCATGACATAGCCTTCATATAAATGGTCTCGCC 125
DB      8688  GGCTGCGAATCTCGACACAATGCTATATGATGATGATTCGTAACAATGGTCTCTCC 8629

QY      126  ACCTACCTGAATGAGGATACGACGCTGCGCTTGCCCTTGCTTGCTGCTGCTGCTG 182
DB      8628  ACAAGCATGATGTAAGCGATAGTGGCTTGACGATCCGACATGATGTTTGTATGGA 8569

QY      183  TTTCAATTTGCCGATCGGTTGAAAATGGAATAGCTATGACTGTAAAAAGTACGAT 242
DB      8568  AAGAGCATCGCTTTCGCGGTTAAAACGGAAGTGAAGTAACTGTGAACAACTCGGGCT 8509

QY      243  TCAATGAGTGATCGTCTGCTAAATAACAATGCGGATCCGAAGCGCATGAAG 302
DB      8508  TTAGTACGCGGACGACGCTTCATGATGCGAATTTCACTTCACTGAAGCGACATCTT 8449

QY      303  CCAACTGATGATTTCTTGCGGCTACTTCTCTAATAATCGCTAATGCTTCACTGCG 362
DB      8448  GAAGTTTAAAGACTTGTGCTTTAAGCTTGTGCGGACATTCGACAGATCTTCTCGCTAC 8389

QY      363  GCAATTCATTCATGACGCACTGCAACATCCGAGATACAAAGATGAAGGCGCTGACA 422
DB      8388  GACGATGAATTTGTATGAACCATGCAATCGAGATTAAGCGCGGACATGCACTATCA 8329

QY      423  TATTGATTTATAGGATCAAGATCCATTAACCTGCGATTAATAAGAGAAATAATAGC 482
DB      8328  TGTTCCTTTGTCACATCAAGTCCATGCACTTGAATGATGACAGAAAGAAATAATAG 8269

QY      483  GATCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 542
DB      8268  TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8209

QY      543  ACGTGTGCTAATCTTGTGCGGCAACATTAAGTACCTTGTAACTTGTCACTTTGTGA 602
DB      8208  ACT-----AATTGCGTAAACATTAACACCCCACTTTAGTTGTCACTTTGTGA 8157

QY      603  CAACTAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 662
DB      8156  CTAAACATCTTTTATATATGTTTTCGCTTAAAAATTCGACTGATGTGATTAAG 8097
```

```

QY      663 TAGCATCAACAACATTAAGTACTCTTGTATCTGATTAAGTTGTCACAAAGTCTTA 722
DB      8096 TCGAGATTATAGGCAATGTTATGCTCAATCGATTTCTGTTACAGAGGTGACAAATTTCT 8037
QY      723 TTACATTGACCAATCATGACACTTACCTTAAATAAGCCGTTGTTA----- 772
DB      8036 AACCACTGCTTACTTACCACTGATTAAGTACCACTTGTATGCGCTGATTAATTCAAAATC 7977
QY      773 ---TAAAGGAAGCCATNTGATTTGTCACATGCAATGATTAATTTGTGCTGCT----- 821
DB      7976 AATATTAGGACCAACCATGATGTAAGTCTGATGATGATTCATTCATTCAGGCTATCTGCT 7917
QY      822 -TGCGCCGAATCTATCCAGGTTTACTGTTGAAACGCTTATCCAAATCGCCAGATT 880
DB      7916 GTGCGGATTAAGGATTTGAAAGTTGCTGATTAAGTTCAATCCGAATGGCTTACC 7857
QY      881 GTGCTTAATGAGGATTTGCTGCGGGATAGTGTATGACGAAGATATGACCGCTCATGTTG 940
DB      7856 ATGGTGTCTGGGCAATTAACGGGCGGGTTGTCTTGTGAATAAGACTTGAACGAGAGCG 7797
QY      941 GGAACCTGTCGATGAGATTTTGAATGACGACGACGATTTTGTGCGCAAAAAGTCC 1000
DB      7796 GACGTCGCGGATGAAACTTTGAAGCGGCGCGCGGACGATTTGCGCGAGAAAGATTC 7737
QY      1001 ATACTTATCTTATTTATGACGATCCGCTGTTGATGCAACCCCAACGCAATCCGA 1060
DB      7736 ATACCTATCTCTGATCACTTCAAGGAGGCTTATGACGAGAGACCTTAAGCGCGATCCGG 7677
QY      1061 ATGCTTGAAGTCTGATTTTCCCATTTAGCTTTTAAACCCGGAATGTCAAACAA 1120
DB      7676 AAGATTGAGCTTAAACATCAACCATTAAGCATTAAGTGAACGCAACATCTGAGCAA 7617
QY      1121 TAGAAGATTTTGTATGACACCCGAGCGGCTTATGTTGATCTTCACTTACTCA 1180
DB      7616 TCAACATGCTGCGCGGCGGAGTGCAGCTCAATATGTTCCCATTTGACGCGATCTCCA 7557
QY      1181 AAGAAAGATGCGCGCTGCTTTTGTATGATGTCGCGCAATTCAGATCGGCGCAAAAT 1240
DB      7556 ATGGAGAAAGAACCTTGAATTTGACCAACAAAAACATCGAAAAACATGGCGGAAAC 7497
QY      1241 TACGCGCTGCGGTTGATGATCAATCAATCCGCTGATCTATTTTCAATTAGAAAAAGTTTATG 1300
DB      7496 TCGGCTGCTTCGATTTAGTACATCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7437
QY      1301 TGGCGGATTTATTTGATGCTTACGCGCAATTTGGCGTGAAGTATATGCAATGACATTA 1360
DB      7436 TCGGAGACATCAATTTCTGATATCAAGAAATTTGGCATGACATGACGCGCATGACATTA 7377
QY      1361 AAGCGCGCTGATCAATACCGGGGATGATGTCATGATCAATTAATTTGCGCGCATTTGTA 1420
DB      7376 ACGCGCGTTTATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7317
QY      1421 AAGGCAAGAGCAACCAACGACCGCTTATGCTTGTGCAATGATGATGATGATGATGATGAT 1480
DB      7316 AAGGCAAGAGCGGATTAAGCAAGCAAGCTGATGATGATGATGATGATGATGATGATGATGAT 7257
QY      1481 TTCAAAACCTGTTTACGAGGTTAAGTGTTCGAAA 1513
DB      7256 TTCAAAACCTGTTTGTGCTGCTTAAAGCAAGCA 7224

```

```

RESULT 5
AP005349      247450 bp   DNA      linear   BCT 04-DEC-2003
LOCUS      Vibrio vulnificus YJ016 DNA, chromosome II, complete genome,
DEFINITION
ACCESSION   AP005349   BA000038
VERSION     AP005349.1   GI:37201329
KEYWORDS
SOURCE      Vibrio vulnificus YJ016
ORGANISM    Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
            Vibrionaceae; Vibrio.

```

```

REFERENCE
AUTHORS      1
              Chen, C.Y., Mu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L.,
              Liu, Y.M., Chen, H.J., Shen, A.B., Li, J.C., Su, T.L., Shao, C.P.,
              Lee, C.T., Hor, L.I., and Tsai, S.P.
              Comparative Genome Analysis of Vibrio vulnificus, a Marine Pathogen
              Genome Res. 13, 2577-2587 (2003)
              2 (bases 1 to 247450)
              Chen, C.Y., Mu, K.M., and Tsai, S.P.
              Direct Submission
              Submitted (29-MAY-2002) Shin Feng Tsai, National Health Research
              Institutes, Division of Molecular and Genomic Medicine, 128,
              Yen-Chin-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China
              (E-mail: pete@nhihri.org.tw, Tel: 886-2-8146-1041,
              Fax: 886-2-2789-0484)
              This sequence was determined by the Sequencing Core of the National
              Yang-Ming University Genome Research Center (YMC;
              http://genome.ym.edu.tw).
              Location/Qualifiers
                1. 247450
                  /organism="Vibrio vulnificus YJ016"
                  /mol_type="genomic DNA"
                  /strain="YJ016"
                  /db_xref="taxon:196600"
                  /chromosome="II"
                  /complement(73..1092)
                  /gene="VVA1126"
                  /complement(73..1092)
                  /gene="VVA1126"
                  /codon_start=1
                  /transl_table=11
                  /product="outer membrane protein N, non-specific porin"
                  /protein_id="BAC97152.1"
                  /db_xref="GI:37201330"
                  /translation="MKKIIALAVASATPQTQAVVEVYNDGTFISIGHVSVALD
                  SKQDGLGTVASPRINFNATODLNGFTADAKGMSINLIDGEMALSTRLVIGLTH
                  ESLGRGVGTQWAPYVAVGADMPLEAFNDIYDQHGMLGDAEEMVYAAPELG
                  EAKGIALGCMGRKTDNNNSGNRAQILANSINLVNVAVNTGDDVYVAGVGKTI
                  SESHVLSATYGGTCKGLIYAGTYANMEYNSGDGMLKRTSAYEPALATLNRGNLS
                  VNYESVIDELSLRVYSTSAVDPEINFTKFPAGYQPDLGSGDYKEDKMWLG
                  ARYL"
                  /complement(1331..1519)
                  /gene="VVA1127"
                  /complement(1331..1519)
                  /gene="VVA1127"
                  /note="identified by Glimmer2"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="BAC97153.1"
                  /db_xref="GI:37201331"
                  /translation="MRAIVYMDISISIKLASRGVIGIRPAFNSPSSLKGESISYCNII
                  INISVFLIDPNVYSVC"
                  /1559..2050
                  /gene="VVA1128"
                  /1559..2050
                  /gene="VVA1128"
                  /codon_start=1
                  /transl_table=11
                  /product="thiol-disulfide isomerase and thioresoxins"
                  /protein_id="BAC97154.1"
                  /db_xref="GI:37201332"
                  /translation="MKIVILKEMLPVLAVLIVSIALDPFYSRNLPKDNALPLSAMT
                  TLGTQVLDLASGSPVVVYFWATWCGPKLVSPITNIDSGYPIVIVAMQSGDEBSV
                  NOYLQSQQYDYDPIINDPSTKISRDMGVQTPVIVIVRNKIEFITSVSPIGLWRLT
                  WLS"
                  /2203..3180
                  /gene="VVA1129"
                  /2203..3180
                  /gene="VVA1129"
                  /codon_start=1
                  /transl_table=11
                  /product="putative outer membrane protein"
                  /protein_id="BAC97155.1"

```

```

/db_xref="GI:37201333"
/translation="MKKALATATILSAVTVGSSPAATVYKDTGTEKIGRYEFRGDF
IGDKGVEIDGTMEDSTRALNVKGTIDLNDQAFYFAEKTGESSKRNMYAG
VKTDAFVSQKDMASVIVSDLDITEFSGVOQFIDASDKTSDTFEAGKSDAIQI
EATASADADNDLMGISGLPIGLDGLASTGNGNGNSOKOILVGAQYTLND
LIGATSTGDLDDKAKOEPFAVQYKISPEFSAAMYTPAEEDNDGTSDAKNG
LELVGYKNSNPTIYLSYLNKLDVQKVDNKGVTSGEDTLRLGVRYDF"
3606..5381
/gene="VVA1130"
3606..5381
/gene="VVA1130"
/translation="MVDGRSKIVANSKVLILKAAYTVAVLITOTLVDBHTLLPSY
VLPAIMIIVALLSPSHIKIVYCALIATSYLFGVIGVDPDIDIVIEETFLILCYLL
PFGTMAFVALLISYFYTPPEGDELEIAELIYVSASIMVYQOKFYIQL
TKREDRTDYTLQPNRAFYSDDNVESLAQGRCALIKIDAFPEINSEFHE
QADLLRFERRLKNTLFGQVYRMEGDFAVLYEKSHINDEVDALLDKRKALTA
AFELERSGYOISVATGALPNESWTNLDIVLRNSDIAAOKATRGQVARYEALD
ETIRHOIERELDAIELKOFSLVYQVPSITTEITGAELIRMOHPELGITTEPEF
ITIAEKTKOIIPIGRWVILBAQOAKKMLGYKIPSVXSVSTQVAFHDDIELQVTA
LAESHLPASYLOLEITETVMEKPKLIEICNLRQEGVTAVIDDGGIAYSSINLYKO
LPIDVLIKDAFIDDCVDNHDHMLVTTIVQLGHNDKXVIAEGVTEEORLLVSEQ
CDEQGYLSPKPDVEEFNRQLQI"
complement (5356..5475)
/gene="VVA1131"
complement (5356..5475)
/gene="VVA1131"
/translation="MSIPKPHNHSKQDNHAKESWCKGNMVGAKAFVATGQ"
5718..6596
/gene="VVA1132"
5718..6596
/gene="VVA1132"
/translation="MOSQHPILGASMLTAGLAFIVNSLAQVASIQPAISTVVAL
IOYGIALLIPLPKTLGIRQSLRSHFGHARVLSVIGLQIMMALAVPVIVOG
IALMTSPFLATIGSGLLEKGVAVRWGTLRFGAMIIIEPMADBDHATLLPVG
AAPPACYSIMWKKLSVNDPSTMVVYLLILTPNIMAIIPFTWPSSTLWILIFA
AGCMTALAOVAIVKAVASADSPVQPFDAKPLANTLAQMLVGVWVPERMLGAFII
IASAVTTHWETKTK"
6661..7266
/gene="VVA1133"
6661..7266
/gene="VVA1133"
/translation="MSEAKITLYRWAGSNPKPKNIPGECTLTNDILDTFANELN
FIVLELVKDMLSHWEPLKLGAMHAPIIVBEKVISQEBALNRGVLISGVKEWTKR
DVLKGINIVFGKATPCYCVKAKOLLDSAGHIGYCHADVQKSAALYRMIPEVKAIIGKTI
PVTVPQIWLDSYIGGCDKLEVLKERGLDVVNNVVEMAN"
complement (7360..7494)
/gene="VVA1134"
complement (7360..7494)
/gene="VVA1134"
/translation="identified by Glimmer2"
/codon_start=1
/codon_start=1

```

```

/translation="identified by Glimmer2"
/codon_start=1
/translation="conserved hypothetical protein"
/product="conserved hypothetical protein"
/protein_id="BAC97156.1"
/db_xref="GI:37201334"
/translation="MVDGRSKIVANSKVLILKAAYTVAVLITOTLVDBHTLLPSY
VLPAIMIIVALLSPSHIKIVYCALIATSYLFGVIGVDPDIDIVIEETFLILCYLL
PFGTMAFVALLISYFYTPPEGDELEIAELIYVSASIMVYQOKFYIQL
TKREDRTDYTLQPNRAFYSDDNVESLAQGRCALIKIDAFPEINSEFHE
QADLLRFERRLKNTLFGQVYRMEGDFAVLYEKSHINDEVDALLDKRKALTA
AFELERSGYOISVATGALPNESWTNLDIVLRNSDIAAOKATRGQVARYEALD
ETIRHOIERELDAIELKOFSLVYQVPSITTEITGAELIRMOHPELGITTEPEF
ITIAEKTKOIIPIGRWVILBAQOAKKMLGYKIPSVXSVSTQVAFHDDIELQVTA
LAESHLPASYLOLEITETVMEKPKLIEICNLRQEGVTAVIDDGGIAYSSINLYKO
LPIDVLIKDAFIDDCVDNHDHMLVTTIVQLGHNDKXVIAEGVTEEORLLVSEQ
CDEQGYLSPKPDVEEFNRQLQI"
complement (5356..5475)
/gene="VVA1131"
complement (5356..5475)
/gene="VVA1131"
/translation="MSIPKPHNHSKQDNHAKESWCKGNMVGAKAFVATGQ"
5718..6596
/gene="VVA1132"
5718..6596
/gene="VVA1132"
/translation="MOSQHPILGASMLTAGLAFIVNSLAQVASIQPAISTVVAL
IOYGIALLIPLPKTLGIRQSLRSHFGHARVLSVIGLQIMMALAVPVIVOG
IALMTSPFLATIGSGLLEKGVAVRWGTLRFGAMIIIEPMADBDHATLLPVG
AAPPACYSIMWKKLSVNDPSTMVVYLLILTPNIMAIIPFTWPSSTLWILIFA
AGCMTALAOVAIVKAVASADSPVQPFDAKPLANTLAQMLVGVWVPERMLGAFII
IASAVTTHWETKTK"
6661..7266
/gene="VVA1133"
6661..7266
/gene="VVA1133"
/translation="MSEAKITLYRWAGSNPKPKNIPGECTLTNDILDTFANELN
FIVLELVKDMLSHWEPLKLGAMHAPIIVBEKVISQEBALNRGVLISGVKEWTKR
DVLKGINIVFGKATPCYCVKAKOLLDSAGHIGYCHADVQKSAALYRMIPEVKAIIGKTI
PVTVPQIWLDSYIGGCDKLEVLKERGLDVVNNVVEMAN"
complement (7360..7494)
/gene="VVA1134"
complement (7360..7494)
/gene="VVA1134"
/translation="identified by Glimmer2"
/codon_start=1
/codon_start=1

```

| Query Match | 8.9% | Score 230.4 | DB 1 | Length 247450 |
|-----------------------|---|-------------------|----------|---------------|
| Best Local Similarity | 62.9% | Pred. No. 1.7e-55 | | |
| Matches 390 | Conservative 0 | Mismatches 226 | Indels 4 | Gaps 2 |
| 1555 | GAAGCAATGCAAGATTATCTTGTTCATGCACTCATATGATGATGGCTTGATATGATCGG | 1614 | | |
| 145723 | GAAGGATGAAGAAATCATATATTAACAAGTTATATATGATGATGGTTAGTCAGCAACT | 145782 | | |
| 1615 | CTTATGATGCTGCGCAAAATTTGGTTATGTAATGTAACCAATAGCTAACAATCACTC | 1674 | | |
| 145783 | CTGAGCTTGCGCTTAATAAATTTGGGCTATCAACAAAGAGCTCACTAACAATCACTC | 145842 | | |
| 1675 | GCTATGATGATGAGGCCATTTTTCGCGCTTGAACCATGCTCACTCATGCTCGCT | 1734 | | |
| 145843 | AATATGATGATGAGAAATTTTTCGCGCTTGAACCATGCTCACTCATGCTCGCT | 145902 | | |
| 1735 | AATGCTTAGTCGACACAGTTTGGGCGAATTTGGTATCAACAGTTATCTAGAAATCGGC | 1794 | | |
| 145903 | AATGCTTAGTCGACACAGTTTGGGCGAATTTGGTATCAACAGTTATCTAGAAATCGGC | 145962 | | |
| 1795 | GCACCGTCTCTGAAACCCCTCTCCCATGCTCGCCCATCGGCTCACTTTGCAAGAGCT | 1854 | | |
| 145963 | CGCCCTACCTTCAACAGATTCATCATGTGTGTGCGATAGAGGTCAACGCTCAAGGCGCT | 146022 | | |
| 1855 | TCCATGTCATTAATTAATTAAGCAATTAAGTTTAAGGGGTGGCACTTAGATTAATTCAGAGAA | 1914 | | |
| 146023 | TCCATGTCATTAATTAATTAAGCAATTAAGTTTAAGGGGTGGCACTTAGATTAATTCAGAGAA | 146082 | | |

QY 1915 TTTCGGTTAAAGACACGACGACGATCCGCTATCCACAAAATCAGCAGTATTGCA 1974
DB 146083 CATGGGCTCAATATGCAATGAGATGTTTGGATTTTCTCTCAAAAGCTTGGAGATTTGCA 146142
QY 1975 GGAACGATACCTTTAGGGCTGCGCAGCCTTTTACTGCCGATCCACTG--GACTCCGAT 2031
DB 146143 GGCACGATTCCTAGTGGGTGCAAGACCTTCTTACTGATGATGATACACACGATGCTGAT 146202
QY 2032 GGTACCGTACAGTGAAGAAACCAAAATAGCTGCGATGACGATGATGCGGATATCC 2091
DB 146203 GGGAGCGTACGCTGAGGAGACAAAGATCTAGGTATGACTGACCACTGTTTGAAGAAA 146262
QY 2092 ACCACTTCATACGAGATGCTGTTTATCATTCCTGTCGCGACCAATGACCACTTCT 2151
DB 146263 AATTA-GCCATACACGATGATATATATAGCCGATATTCGATGATGATGATGATTCAT 146321
QY 2152 TCGTATGACCGCTTCCGCG 2171
DB 146322 CCAGCATTCCTCACTTCAGGC 146341

RESULT 6
AE016810 302331 bp DNA linear BCT 29-SEP-2003
LOCUS VIBRIO vulnificus CMCP6 chromosome II section 3 of 6 of the
DEFINITION Complete sequence.
ACCESSION AE016810 AE016796
VERSION AE016810.1 GI:27358548
KEYWORDS
SOURCE
ORGANISM
Vibrio vulnificus CMCP6
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrio vulnificus CMCP6

REFERENCE
AUTHORS
Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, B.K., Shin, D.H.,
1 (bases 1 to 302331)
Chung, S.S., Choy, H.E., Proguileke-Fox, A., Hillman, J.D., Handfield, M.
and Rhee, J.H.
Characterization and Pathogenic Significance of Vibrio vulnificus
Antigens Preferentially Expressed in Septicemic Patients
Infect. Immun. 71 (10), 5461-5471 (2003)

TITLE
JOURNAL
PUBMED
14500463
2 (bases 1 to 302331)
Jeong, H., Moon, Y.H. and Kim, J.J.
Direct Submission
Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
Yuseong-gu, Daejeon 305-811, South Korea
3 (bases 1 to 302331)
Rhee, J.H., Kim, S.Y., Chung, S.S., Lee, S.E. and Choy, H.E.
Direct Submission
Submitted (13-DEC-2002) Department of Microbiology, Genome Research
Center for Enteropathogenic Bacteria, Chonnam National University
Medical School, Hak-Ji-Dong, Dong-gu, Kwang-Ju 501-746, South Korea
Location/Qualifiers

FEATURES
source
1.302331
/organism="Vibrio vulnificus CMCP6"
/mol_type="genomic DNA"
/strain="CMCP6"
/db_xref="taxon:216895"
/chromosome="II"
complement(70..1368)
/locus_tag="VV20549"
complement(70..1368)
/locus_tag="VV20549"
/note="COG325"
/codon_start=1
/transl_table=11
/product="Chitinase"
/protein_id="AA007498.1"
/db_xref="GI:27358549"
/translation="MKKTLITAFSAFAGSVAIOEKVAVGYPADMOYANASPYTVK
DPAKLTIVIAAPLSMCPHTGASETVOKLVAKCGKEPESALIVDTALAKDGG
KSVKVPFKGHAQLAELKQKPDKIIPSGRTMSRPFPAKADPAMAKQPKTAV
ELIKQYDFPDIDIMEYIPGGGLITSPNPATKUSDQKAERDAFTYLVKTLRADL

DALEKTNREVELSTAVGAGAAQIDMKAAQPYLTNNFAMTYDELAGGMOQTGHTTN
LHATERSWNGADYVFINQIELGIPSEKLVGAAPYGGMGQGTGDDQAKGDLIS
EGAGRGNGENYFMWMDLVKNYGRKQGVAVYDQSOAPRYLMBKRYFISFEDQRS
IRAKQMAQSNLGGIFTLELSDGSGKLVDMHLEMTSK"
complement(1510..2502)
/locus_tag="VV20550"
complement(1510..2502)
/locus_tag="VV20550"
/note="contains sigma factor-related N-terminal domain;
COG2390"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AA007499.1"
/db_xref="GI:27358550"
/translation="MSSMODISISTDLITLISVAVYDGAFTOEISKPSVRAKV
GRILKQARDGIVEITVKYHPYPSAKIBERLLERGVRAALVLAQHPHQAQVAG
LVSNVLTSLKGVVSVGGQNVSAHVAHVITSRDQFVGGITPRGATRAD
HICROLAKRYGQTSBTLVAPAYENREQMAQMAVQTLIDARKADILIGIDM
SENSVYVDLGMFTPEDEVQSRLOQGVDFAGYDFEDIGANTVMTSDRVYGLGIE
FRISQVIAIAENSKPLALGALRTGADIVATSVSNALTVINIDECQGRAQ"
2966..3916
/locus_tag="VV20552"
2966..3916
/locus_tag="VV20552"
/note="COG0176"
/codon_start=1
/transl_table=11
/product="Transaldolase"
/protein_id="AA007500.1"
/db_xref="GI:27358551"
/translation="MSNKLQRLKLTTVADTGEIDAIKKYQPEDATTPSLIKRAQ
IAEAPLIDASILEYATQSDDKAQOIQDCLMAVITGKELIKTIPRISTEDALIS
YMEGSAVARAIEVATYNDAGITNDRILIKLASWEGIRPAALIEKEGINCLTLLFS
PAQARACAGVELISPFVGRIMDYKAGEGDFEAQEDPGLSVTKIINYYKEGYK
TVYMGASFNIGIIEIACDRLLTAPALAEBAAGEVEKLVDSKGAARBPAMT
HAEFLMEHNDPMVAEKLAEGRINFAVDQKLEMIATL"

REFERENCE
AUTHORS
Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, B.K., Shin, D.H.,
1 (bases 1 to 302331)
Chung, S.S., Choy, H.E., Proguileke-Fox, A., Hillman, J.D., Handfield, M.
and Rhee, J.H.
Characterization and Pathogenic Significance of Vibrio vulnificus
Antigens Preferentially Expressed in Septicemic Patients
Infect. Immun. 71 (10), 5461-5471 (2003)

TITLE
JOURNAL
PUBMED
14500463
2 (bases 1 to 302331)
Jeong, H., Moon, Y.H. and Kim, J.J.
Direct Submission
Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
Yuseong-gu, Daejeon 305-811, South Korea
3 (bases 1 to 302331)
Rhee, J.H., Kim, S.Y., Chung, S.S., Lee, S.E. and Choy, H.E.
Direct Submission
Submitted (13-DEC-2002) Department of Microbiology, Genome Research
Center for Enteropathogenic Bacteria, Chonnam National University
Medical School, Hak-Ji-Dong, Dong-gu, Kwang-Ju 501-746, South Korea
Location/Qualifiers

FEATURES
source
1.302331
/organism="Vibrio vulnificus CMCP6"
/mol_type="genomic DNA"
/strain="CMCP6"
/db_xref="taxon:216895"
/chromosome="II"
complement(70..1368)
/locus_tag="VV20549"
complement(70..1368)
/locus_tag="VV20549"
/note="COG325"
/codon_start=1
/transl_table=11
/product="Chitinase"
/protein_id="AA007498.1"
/db_xref="GI:27358549"
/translation="MKKTLITAFSAFAGSVAIOEKVAVGYPADMOYANASPYTVK
DPAKLTIVIAAPLSMCPHTGASETVOKLVAKCGKEPESALIVDTALAKDGG
KSVKVPFKGHAQLAELKQKPDKIIPSGRTMSRPFPAKADPAMAKQPKTAV
ELIKQYDFPDIDIMEYIPGGGLITSPNPATKUSDQKAERDAFTYLVKTLRADL

DALEKTNREVELSTAVGAGAAQIDMKAAQPYLTNNFAMTYDELAGGMOQTGHTTN
LHATERSWNGADYVFINQIELGIPSEKLVGAAPYGGMGQGTGDDQAKGDLIS
EGAGRGNGENYFMWMDLVKNYGRKQGVAVYDQSOAPRYLMBKRYFISFEDQRS
IRAKQMAQSNLGGIFTLELSDGSGKLVDMHLEMTSK"
complement(1510..2502)
/locus_tag="VV20550"
complement(1510..2502)
/locus_tag="VV20550"
/note="contains sigma factor-related N-terminal domain;
COG2390"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AA007499.1"
/db_xref="GI:27358550"
/translation="MSSMODISISTDLITLISVAVYDGAFTOEISKPSVRAKV
GRILKQARDGIVEITVKYHPYPSAKIBERLLERGVRAALVLAQHPHQAQVAG
LVSNVLTSLKGVVSVGGQNVSAHVAHVITSRDQFVGGITPRGATRAD
HICROLAKRYGQTSBTLVAPAYENREQMAQMAVQTLIDARKADILIGIDM
SENSVYVDLGMFTPEDEVQSRLOQGVDFAGYDFEDIGANTVMTSDRVYGLGIE
FRISQVIAIAENSKPLALGALRTGADIVATSVSNALTVINIDECQGRAQ"
2966..3916
/locus_tag="VV20552"
2966..3916
/locus_tag="VV20552"
/note="COG0176"
/codon_start=1
/transl_table=11
/product="Transaldolase"
/protein_id="AA007500.1"
/db_xref="GI:27358551"
/translation="MSNKLQRLKLTTVADTGEIDAIKKYQPEDATTPSLIKRAQ
IAEAPLIDASILEYATQSDDKAQOIQDCLMAVITGKELIKTIPRISTEDALIS
YMEGSAVARAIEVATYNDAGITNDRILIKLASWEGIRPAALIEKEGINCLTLLFS
PAQARACAGVELISPFVGRIMDYKAGEGDFEAQEDPGLSVTKIINYYKEGYK
TVYMGASFNIGIIEIACDRLLTAPALAEBAAGEVEKLVDSKGAARBPAMT
HAEFLMEHNDPMVAEKLAEGRINFAVDQKLEMIATL"

CDS

1/locus_tag="VW20556"
6861..9254
/locus_tag="VW20556"
/note="other COG candidate exists; contains a membrane domain, an EAL and a GDEF domain; COG5001"
/codon_start=1
/transl_table=11
/product="Predicted signal transduction protein"
/protein_id="AAO07503.1"
/db_xref="GI:27358554"

translation="MNSLFGHYLRISRRFTFLALISIQVSVLVVTVDAHI
REIRLEAVSRVWDQLIQASFRLOQSNADLADFEPRVAAPDQDTTSLQNN
AARVADIASLIDHOMALQALSEDDISESRKELVATLKQSTAPYHLMYKRP
FYALAPRIRIGVLFGEITRHSIASQIAGDIALISLSISNIIYVSEAO
KAFPSLAEIRPTASTMODRALIDRVETETELPNOHMIVVYFESLDAIAP
DKLASOYIMVSGIAIVLTALALSHRIVRLITMTKATSLDAGVATTEGQRP
DEIGLOTRGFDOMERLEQKQTTTQIAFPDITLTPRHRFRQKRELSTKSI
SVITINDRFKINDVIGHVGDVAVLQEMAQRLITSSLTLPARISGDFVEIKT
GSDLEADSVTRIKORLIDIPVNDPDIISVIGVATPSSGDGDVSLNLSQIAMS
ARKKEQVFPYDSALVTAIPEHLSLSLRLALONELRVLPQVMAVTVVSAEA
LIRMOHPERGLAPFPVPAQOTGYVELTRMMIQOYIOHRELQPLSGQLTISVL
STWDLINPOLDLLOHMLSEYHVPASGCLLETSSALIMEDETSLLTLOLSOMFHL
STDDRCGSSSGVQLRPLVNEKLKDKSFLVAMLENPNDRKIVSTTHLSLGLQVY
AEGVSYDLFTSLRELQCDRIQICVSKPLALDEFTIAWEEWOSNPYVEMOKG"

gene

1/locus_tag="VW20557"
complement(9304..10866)
/locus_tag="VW20557"
/locus_tag="VW20557"
/note="other COG candidate exists; COG0840"
/codon_start=1
/transl_table=11
/product="MetH1-accepting chemotaxis protein"
/protein_id="AAO07504.1"
/db_xref="GI:27358555"

translation="MSALISNSAHETLLKEHOLVSTTDLKGVTTCNEAFRTAIFS
RQELIGENHNVIRSHDMKRAFGMOMLKGCKMGTGVRKRTSGGWDVATVPI
YENGWMSGYQSVKPKREKRWYIASKAHSLIDERNRQMSLKNLSIRATILGSI
SAPVLAQMLSDGTSTLANWPLTTLALFRQELIDTPQQLKALQSEPSVLITS
GKNQSIADFHIRKASARIKVLGEMTSARPLQADLDELTAKEVSOALSQASLI
RQVQATDSMEVMAQSVSSIQDADRLEETFORVAKQSIIDTHRELMQIAEK
ATATYQLSBOAKQSVFMEIGGIAEOTNLALAAAEARAGQSGGVVADVRA
ALSGRNTATLOIRSIDTMTLATIOGMDLLENKQCESEVAALSAELSVBOM
MRSQMLATYVLAADNOMQLSSQVNHHSIVTTSQNLAATYSVENSKLSKSRD
DPYGLAYRFEK"

gene

1/locus_tag="VW20558"
complement(11177..12229)
/locus_tag="VW20558"
/locus_tag="VW20558"
/note="DAH1; COG0722"
/codon_start=1
/transl_table=11
/product="3-deoxy-D-arabino-heptulosonate 7-phosphate synthase"
/protein_id="AAO07505.1"
/db_xref="GI:27358556"

translation="MRTQDVIRINNVRKELLPVAVLEKYPATETASSTFOSRDINH
ILIGDDELIVYGCSDHDPVRAVEYKRLATDELKGLETVMVYFEPKPTTG
WKLGLNDPLNDTYKLDGKLRKGLLDLDLDMGEPHSEPLDITPQYADLSKGA
IGARTGSDVHRELASGLSCVPFGNIGDINKIASDILRASASHHLSKXGSHA
IVERTGNDPCHILIRGKEPNVAHVAVKELATGLPKVWIDESHANSQOYR
QMLVADVSECLAGEBAIFGVMIHSHLVESRQDVIDKATATGCSITDAGMEDTE
TVLRQLANAVABRAK"

gene

1/locus_tag="VW20559"
12652..13725
/locus_tag="VW20559"
/locus_tag="VW20559"
/note="peptidoglycan-associated (lipo) proteins; COG2885"
/codon_start=1
/transl_table=11
/product="Outer membrane protein"
/protein_id="AAO07506.1"
/db_xref="GI:27358557"

translation="MKKLAVTLASVLAASSGFPSSATANAAMSDLYFGKIGYVNLDDAC

YVSPQDDSDAGFLVYGVNFSKHISAEIGDILGHEHNSNGYSEAKLAAYTLAP
KPSFLPANEKIDAPAKIGAAIMTYGDDKDLVPTSLGLFVLTLSFPARAEYQYQDMS
DDIVKMDANFPGIGITLYLRGASASATVAAAANVAVESRPAEVRKQAEVAFILL
AEEPKEPKEWTKPSNOQYNQELFATGSSKLISAGKQKLTPLAEVLIKYPAEAEI

Query Match 8.9%; Score 230.4; DB 1; Length 302331;
Best Local Similarity 62.9%; Pred. No. 1.8e-55;
Matches 390; Conservative 0; Mismatches 226; Indels 4; Gaps 2;

QY 1555 GAGCAATGCAATATTCTTGTTCATGCACTCATATGATGGCTTGGTAATGATCCG 1614
Db 259431 GAGCCGATGAAATCATATATTAACACGGTTTATATGATGATGATGATGCAACCT 259490
QY 1615 CTTAGCATGCTGCTGCAATAATGGGTATCGTACCAACCAATTGGCTCAACTGCTC 1674
Db 259491 CTTAGCTTGGCTCTTAAAAAATGGGTATGCAACACCAACGCTCACTTCAACACGGCTC 259550
QY 1675 GCTATCGATGATGAGCCATTTTTCGCGCGCTTGAACGATGCTCATCTATGCTCGCT 1734
Db 259551 AATATGATGAGAAAAAGTTTTCATATGATGATGATGATGATGATGATGATGATGATGAT 259610
QY 1735 AATGCTTTAGTCGACACAGCTTTGGCGGATTTGGTATGCAACGTTATCTAGATGCGCG 1794
Db 259611 AATGCTTTAGTCGACACAGCTTTGGCGGATTTGGTATGCAACGTTATCTAGATGCGCG 259670
QY 1795 GCAACCGCTCTGTGAAACCCCTCTCCATGTCGTCGCGCATGCGCTCACTTTGCAAGAGCT 1854
Db 259671 CGGCTTACCTCTTCAAGCAAACTCATATGATGATGATGATGATGATGATGATGATGATGAT 259730
QY 1855 TCCATTTGCTAATAAATTTAGCAATTAAGTTTGGGTGCGCACTAGTAAATTCAGAGAA 1914
Db 259731 TCCATTTGCTAATAAATTTAGCAATTAAGTTTGGGTGCGCACTAGTAAATTCAGAGAA 259790
QY 1915 TTGGGTTAAAGAAACAGACGACGAAATCCGCTATCCACAAATCAGGAGATTTGCA 1974
Db 259791 CATGGGCTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 259850
QY 1975 GAAACGATACCTTTAAGGCTGCGACGCTTTTACTGCGCGATCACTG--GACTCCGAT 2031
Db 259851 GGAACGATACCTTTAAGGCTGCGACGCTTTTACTGCGCGATCACTG--GACTCCGAT 259910
QY 2032 GGTACGCTCACTAGAGAAACCAAAATAGCTGGGATGACATATCGGATATCC 2091
Db 259911 GGTACGCTCACTAGAGAGAAACCAAAATAGCTGGGATGACATATCGGATATCC 259970
QY 2092 ACCACTTCATACAGATGCTGTTTATCATCTCCGTCGCGAGCAATGACCACTTCT 2151
Db 259971 AATATGCAATACAGATGATATATAGCGGCTATATTCAGATCAGATCAGATCAGATCAGAT 260029
QY 2152 TCGTTATGACCGCTTCCGCG 2171
Db 260030 CAGACATTCATCATTGACGCG 260049

RESULT 7
AX023555/c AX023555 4928 bp DNA linear PART 15-SEP-2000
LOCUS Sequence 10 from Patent WO0003008.
DEFINITION AX023555
ACCESSION AX023555
VERSION AX023555.1 GI:10183916
KEYWORDS
SOURCE
ORGANISM
artificial sequences.
REFERENCE
1 Barth, G., Juretzek, T. and Mauerberger, S.
AUTHORS Recombinant haploid or diploid yarrowia lipolytica cells for the
TITLE functional heterologous expression of cytochrome p450 systems
JOURNAL Patent: WO 0003008-A 10 20-JAN-2000;
UNIV DRESDEN TECH (DE)
FEATURES
SOURCE location/Qualifiers
1..4928
/organism="synthetic construct"

ORIGIN

Query Match 8.8%; Score 226.4; DB 6; Length 4928;
Best Local Similarity 83.4%; Pred. No. 1.3e-54;
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LTR

2271 ATGAGAGCTTCCCATGCAATTAACAATCCGCGACTTTACGCTGCGCGCTTAATAA 2330
2483 ATTAGAGTTACCTGTCGAGTACTTAATCTCCAGAGCGAGTGTACATGAAATTAT 2424
2331 TTGGCAAGTGTCTGCGCGATACGCTGATGCGCATATGTTAAGCCAGCCCGACACCGC 2390
2423 GGTGCACTCTCAGTACATCTGCTGATGCGCGCATATGTTAAGCCAGCCCGACACCGC 2364
2391 CAACACCCGCTGACGCGCCCTGATGCGGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2450
2363 CAACACCCGCTGACGCGCCCTGATGCGGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2304
2451 CTGTGACCGCTCTCGGAGGCTGATGTCAGAGTTTCAACGTCATCAACGAAACGG 2510
2303 CTGTGACCGCTCTCGGAGGCTGATGTCAGAGTTTCAACGTCATCAACGAAACGG 2244
2511 CGAGACGAAAGGCGCTGATGTCAGAGTTTCAACGTCATCAACGAAATATG 2570
2243 CGAGACGAAAGGCGCTGATGTCAGAGTTTCAACGTCATCAACGAAATATG 2184
2571 TTCTTAG 2578
2183 TTCTTAG 2176

RESULT 8
AX023554/c 5391 bp DNA linear PAT 15-SEP-2000
LOCUS AX023554
DEFINITION Sequence 9 from Patent WO0003008.
ACCESSION AX023554
VERSION AX023554.1 GI:10183915
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Barth, G., Juretzek, T. and Mauersberger, S.
TITLE Recombinant haploid or diploid yarrowia lipolytica cells for the functional heterologous expression of cytochrome p450 systems
JOURNAL Patent: WO 0003008-A 9 20-JAN-2000;
UNIV DRESDEN TECH (DE)
FEATURES
Source Location/Qualifiers
1. 5391
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="ringf rmg geschlossene DOUBLestr ngige DNA, Yarrowia lipolytica/Bescherichia coli-DNA Plasmid p67zb"
107. 456
/note="Zeta-Element des Retrotransposons aus Yarrowia lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174, 2477-82"
complement(2904. 3634)
/note="Zeta-Element des Retrotransposons Y11 aus Yarrowia lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174, 2477-82"

ORIGIN

Query Match 8.8%; Score 226.4; DB 6; Length 5391;

Best Local Similarity 83.4%; Pred. No. 1.3e-54;
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LTR

2271 ATGAGAGCTTCCCATGCAATTAACAATCCGCGACTTTACGCTGCGCGCTTAATAA 2330
2342 ATTAGAGTTACCTGTCGAGTACTTAATCTCCAGAGCGAGTGTACATGAAATTAT 2883
2331 TTGGCAAGTGTCTGCGCGATACGCTGATGCGCATATGTTAAGCCAGCCCGACACCGC 2390
2882 GGTGCACTCTCAGTACATCTGCTGATGCGCGCATATGTTAAGCCAGCCCGACACCGC 2823
2391 CAACACCCGCTGACGCGCCCTGATGCGGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2450
2822 CAACACCCGCTGACGCGCCCTGATGCGGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2763
2451 CTGTGACCGCTCTCGGAGGCTGATGTCAGAGTTTCAACGTCATCAACGAAACGG 2510
2762 CTGTGACCGCTCTCGGAGGCTGATGTCAGAGTTTCAACGTCATCAACGAAACGG 2703
2511 CGAGACGAAAGGCGCTGATGTCAGAGTTTCAACGTCATCAACGAAATATG 2570
2702 CGAGACGAAAGGCGCTGATGTCAGAGTTTCAACGTCATCAACGAAATATG 2643
2571 TTCTTAG 2578
2642 TTCTTAG 2635

RESULT 9
AX023552/c 7057 bp DNA linear PAT 15-SEP-2000
LOCUS AX023552
DEFINITION Sequence 7 from Patent WO0003008.
ACCESSION AX023552
VERSION AX023552.1 GI:10183913
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Barth, G., Juretzek, T. and Mauersberger, S.
TITLE Recombinant haploid or diploid yarrowia lipolytica cells for the functional heterologous expression of cytochrome p450 systems
JOURNAL Patent: WO 0003008-A 7 20-JAN-2000;
UNIV DRESDEN TECH (DE)
FEATURES
Source Location/Qualifiers
1. 7057
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="ringf rmg geschlossene DOUBLestr ngige DNA, Yarrowia lipolytica/Bescherichia coli-DNA Plasmid p67ICDPro"
complement(16. 2151)
/note="Promotor des ICU1-Gens aus Yarrowia lipolytica EMBL-Datenbank X=72848"
4654. 5384
/note="Zeta-Element des Retrotransposon Y11 aus Yarrowia lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174, 2477-82"

ORIGIN

Query Match 8.8%; Score 226.4; DB 6; Length 7057;
Best Local Similarity 83.4%; Pred. No. 1.4e-54;
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LTR

2271 ATGAGAGCTTCCCATGCAATTAACAATCCGCGACTTTACGCTGCGCGCTTAATAA 2330
4692 ATTAGAGTTACCTGTCGAGTACTTAATCTCCAGAGCGAGTGTACATGAAATTAT 4633
2331 TTGGCAAGTGTCTGCGCGATACGCTGATGCGCATATGTTAAGCCAGCCCGACACCGC 2390
4632 GGTGCACTCTCAGTACATCTGCTGATGCGCGCATATGTTAAGCCAGCCCGACACCGC 4573

QY 2391 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAG 2450
DB 4572 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTCTCCCGCATCCGCTTACAGACAG 4513
QY 2451 CTGTGACCGCTCCCGGAGCTGATGTGCAGAGGTTTCAACCGCATCACCGAAACGCG 2510
DB 4512 CTGTGACCGCTCCCGGAGCTGATGTGCAGAGGTTTCAACCGCATCACCGAAACGCG 4453
QY 2511 CGAGACGAAAGGGCCCTGATACGCTATTTTATAGGTTATGATGATATATATG 2570
DB 4452 CGAGACGAAAGGGCCCTGATACGCTATTTTATAGGTTATGATGATATATATG 4393
QY 2571 TTTCTTAG 2578
DB 4392 TTTCTTAG 4385

RESULT 10
AX023551/c
LOCUS AX023551 9402 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 6 from Patent WO0003008.
ACCESSION AX023551
VERSION AX023551.1 GI:10183912
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE
1 Barth, G., Juretzek, T. and Mauersberger, S.
AUTHORS Recombinant haploid or diploid yarrowia lipolytica cells for the
TITLE functional heterologous expression of cytochrome p450 systems
JOURNAL Patent: WO 0003008-A 6 20-JAN-2000;
UNIV DRESDEN TECH (DE)

FEATURES
Source Location/Qualifiers
1..9402
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="ringf rmg geschlossene DOUBLER ngise DNA,
Yarrowia lipolytica/Escherichia coli-DNA Plasmid p671C17a"
terminator complement(277..9373)
/note="Terminator des ICL1-Gens aus Yarrowia lipolytica"
intron complement(1959..2354)
/note="Intron in ICL1-Gen aus Yarrowia lipolytica
EMBL-Datenbank X=72848"
promoter complement(2361..4496)
/note="Promotor des ICL1-Gens aus Yarrowia lipolytica
EMBL-Datenbank X=72848"
LTR 6999..7799
/note="Zeta-Element aus LTR des Retrotransposon Y11 aus
Yarrowia lipolytica. Schmidt-Berger et al. 1994, J
Bacteriol 176,2477-82"

ORIGIN
Query Match 8.8%; Score 226.4; DB 6; Length 9402;
Best Local Similarity 83.4%; Pred. No. 1.4e-54;
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 2271 ATCAGAGCTTCCCATGACATTAACATCCGACCTTACGTCGCGCTTAACTAA 2330
DB 7037 ATTAGAGTTACCTGTGCGATGATCACTCTCAAGAGATGATACATGAAATTAT 6978
QY 2331 TTGGCAAGTGTCTGCCGCGATAGCGTATGCGCATAGTTAAGCAGCCCGACACCGC 2390
DB 6977 GGTGACCTCTCAGTCAATCTGCTGATGCGCATAGTTAAGCAGCCCGACACCGC 6918
QY 2391 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTCTCCCGCATCCGCTTACAGACAG 2450
DB 6917 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTCTCCCGCATCCGCTTACAGACAG 6858
QY 2451 CTGTGACCGCTCCCGGAGCTGATGTGCAGAGGTTTCAACCGCATCACCGAAACGCG 2510
DB 6857 CTGTGACCGCTCCCGGAGCTGATGTGCAGAGGTTTCAACCGCATCACCGAAACGCG 6798

QY 2511 CGAGACGAAAGGGCCCTGATACGCTATTTTATAGGTTATGATGATATATATG 2570
DB 6797 CGAGACGAAAGGGCCCTGATACGCTATTTTATAGGTTATGATGATATATATG 6738
QY 2571 TTTCTTAG 2578
DB 6737 TTTCTTAG 6730

RESULT 11
AX023553/c
LOCUS AX023553 10890 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 8 from Patent WO0003008.
ACCESSION AX023553
VERSION AX023553.1 GI:10183914
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE
1 Barth, G., Juretzek, T. and Mauersberger, S.
AUTHORS Recombinant haploid or diploid yarrowia lipolytica cells for the
TITLE functional heterologous expression of cytochrome p450 systems
JOURNAL Patent: WO 0003008-A 8 20-JAN-2000;
UNIV DRESDEN TECH (DE)

FEATURES
Source Location/Qualifiers
1..10890
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="ringf rmg geschlossene DOUBLER ngise DNA,
Yarrowia lipolytica/Escherichia coli-DNA Plasmid p671L43"
terminator complement(3..274)
/note="Terminator des ICL1-Gens aus Yarrowia lipolytica"
intron complement(3475..3801)
/note="Intron in ICL1-Gen aus Yarrowia lipolytica
EMBL-Datenbank X=72848"
promoter complement(3801..5983)
/note="Promotor des ICL1-Gens aus Yarrowia lipolytica
EMBL-Datenbank X=72848"
LTR 8486..9216
/note="Zeta-Element des Retrotransposons Y11 aus Yarrowia
lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174,
2477-82"

ORIGIN
Query Match 8.8%; Score 226.4; DB 6; Length 10890;
Best Local Similarity 83.4%; Pred. No. 1.5e-54;
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 2271 ATCAGAGCTTCCCATGACATTAACATCCGACCTTACGTCGCGCTTAACTAA 2330
DB 8524 ATTAGAGTTACCTGTGCGATGATCACTCTCAAGAGATGATACATGAAATTAT 8465
QY 2331 TTGGCAAGTGTCTGCCGCGATAGCGTATGCGCATAGTTAAGCAGCCCGACACCGC 2390
DB 8464 GGTGACCTCTCAGTCAATCTGCTGATGCGCATAGTTAAGCAGCCCGACACCGC 8405
QY 2391 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTCTCCCGCATCCGCTTACAGACAG 2450
DB 8404 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTCTCCCGCATCCGCTTACAGACAG 8345
QY 2451 CTGTGACCGCTCCCGGAGCTGATGTGCAGAGGTTTCAACCGCATCACCGAAACGCG 2510
DB 8344 CTGTGACCGCTCCCGGAGCTGATGTGCAGAGGTTTCAACCGCATCACCGAAACGCG 8285
QY 2511 CGAGACGAAAGGGCCCTGATACGCTATTTTATAGGTTATGATGATATATATG 2570
DB 8284 CGAGACGAAAGGGCCCTGATACGCTATTTTATAGGTTATGATGATATATATG 8225
QY 2571 TTTCTTAG 2578

Db 8224 TTCTTAG 8217

RESULT 12
AR371489/c
LOCUS AR371489 1073 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 6 from patent US 6395485.
ACCESSION AR371489
VERSION AR371489.1 GI:34608427
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1073)
De Beuckeleer, M.
AUTHORS Methods and kits for identifying elite event gat-zml in biological samples
TITLE
JOURNAL Patent: US 6395485-A 6 28-MAY-2002;
FEATURES
Location/Qualifiers
source 1..1073
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 8.7%; Score 224.8; DB 6; Length 1073;
Best Local Similarity 99.1%; Pred. No. 2.9e-54;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 2410
DB 709 TGCTTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 650
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGACCGTCCGGAGC 2470
DB 649 TGAAGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGACCGTCCGGAGC 590
QY 2471 TGACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 2530
DB 589 TGACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 530
QY 2531 ATACGCTTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578
DB 529 ATACGCTTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 482

RESULT 13
AX195443/c
LOCUS AX195443 1073 bp DNA linear PAT 28-AUG-2001
DEFINITION Sequence 6 from Patent WO0151654.
ACCESSION AX195443
VERSION AX195443.1 GI:15385989
KEYWORDS
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE
1
de Beuckeleer, M.
AUTHORS Methods and kits for identifying elite event gat-zml in biological samples
TITLE
JOURNAL Patent: WO 0151654-A 6 19-JUL-2001;
FEATURES
Location/Qualifiers
source 1..1073
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="sequence comprising a 5' flanking region of
GAT-ZML"

misc_feature 1..341
/note="plant DNA"
primer_bind 286..307
/note="binding site of primer COR17"

misc_feature 342..1073
/note="T-DNA of pUC/Ac"
primer_bind complement(466..487)
/note="binding site of primer COR18"

ORIGIN
Query Match 8.7%; Score 224.8; DB 6; Length 1073;
Best Local Similarity 99.1%; Pred. No. 2.9e-54;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 2410
DB 709 TGCTTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 650
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGACCGTCCGGAGC 2470
DB 649 TGAAGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGACCGTCCGGAGC 590
QY 2471 TGACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 2530
DB 589 TGACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 530
QY 2531 ATACGCTTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578
DB 529 ATACGCTTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 482

RESULT 14
BD261822
LOCUS BD261822 2297 bp DNA linear PAT 17-JUL-2003
DEFINITION Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof.
ACCESSION BD261822
VERSION BD261822.1 GI:33071590
KEYWORDS JP 2002533126-A/11.
SOURCE
ORGANISM synthetic construct
REFERENCE
1 (bases 1 to 2297)
Bestetti, G., Galli, S., Ghisotti, D., Orsini, G., Tonon, G. and Zuffi, G.
AUTHORS Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
TITLE
JOURNAL Patent: JP 2002533126-A 11 08-OCT-2002;
COMMENT
NORPHARMA SPA
OS Artificial Sequence
PN JP 2002533126-A/11
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
ORSINI,
PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: cloning vector derived CC
from pUC18
FH key Location/Qualifiers
FT source 1..2297
/organism="Artificial Sequence".
FEATURES
Location/Qualifiers
source 1..2297
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Query Match 8.7%; Score 224.8; DB 6; Length 2297;
Best Local Similarity 99.1%; Pred. No. 3.4e-54;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 2410
DB 133 TGCTTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCTGACGGCC 192

| | | | |
|----|------|---|------|
| QY | 2411 | TGACGGGCTTTCCTCCTCCCGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGAGGC | 2470 |
| D6 | 193 | TGACGGGCTTTCCTCCTCCCGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGAGGC | 252 |
| QY | 2471 | TGCATGCTCAGAGGTTTTCAACCGTCATCCAGACCGGAGACGAAAGAGGCTCGTG | 2530 |
| D6 | 253 | TGCATGCTCAGAGGTTTTCAACCGTCATCCAGACCGGAGACGAAAGAGGCTCGTG | 312 |
| QY | 2531 | ATAAGCCATATTTTATAGTAAATGCTATGATTAATAAATAGTCTTCTAG | 2578 |
| D6 | 313 | ATAAGCCATATTTTATAGTAAATGCTATGATTAATAAATAGTCTTCTAG | 360 |

RESULT 15
AV037919

| LOCUS | AX027819 | 2297 bp | DNA | linear | PAT 16-SEP-2000 |
|------------|------------------------------------|-------------|-----|--------|-----------------|
| DEFINITION | Sequence 11 from Patent WO0039307. | | | | |
| ACCESSION | AX027819 | | | | |
| VERSION | AX027819.1 | GI:10188663 | | | |

| SOURCE ORGANISM | synthetic construct | synthetic construct |
|-----------------|---------------------|---------------------|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |
| 10 | 11 | 12 |
| 13 | 14 | 15 |
| 16 | 17 | 18 |
| 19 | 20 | 21 |
| 22 | 23 | 24 |
| 25 | 26 | 27 |
| 28 | 29 | 30 |
| 31 | 32 | 33 |
| 34 | 35 | 36 |
| 37 | 38 | 39 |
| 40 | 41 | 42 |
| 43 | 44 | 45 |
| 46 | 47 | 48 |
| 49 | 50 | 51 |
| 52 | 53 | 54 |
| 55 | 56 | 57 |
| 58 | 59 | 60 |
| 61 | 62 | 63 |
| 64 | 65 | 66 |
| 67 | 68 | 69 |
| 70 | 71 | 72 |
| 73 | 74 | 75 |
| 76 | 77 | 78 |
| 79 | 80 | 81 |
| 82 | 83 | 84 |
| 85 | 86 | 87 |
| 88 | 89 | 90 |
| 91 | 92 | 93 |
| 94 | 95 | 96 |
| 97 | 98 | 99 |
| 100 | 101 | 102 |

REFERENCE

| TITLE | AUTHORS |
|---|--|
| Recombinant bacterial strains for the production of natural | Besterlet, G., Call, S., Orsini, G., Tonon, G., Zilli, G. and Ghibetti, D. |

JOURNAL

ESIZIINI GIUSEPPINA (IT) ; CALLI SIMONA (IT) ; NORTHBACON SFA (IT) ;
 OSSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
 GHISOTTI DANIELA (IT) ;

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .2297 |

```

/organism="synthetic construct"
/mol_type="unassigned DNA"
/seq_id="b1000000000"

```

```

/aa_xref="taxon:32630"
/note="cloning vector derived from pUC18"

```

ORIGIN

Query Match 8.7%; Score 224.8; DB 6; Length 2297;

```

      BEB LOCAL SIMILARITY    99.1%    FREQ. NO. 3.4E-54
Matches  226; Conservative    0; Mismatches    2; Indels    0; Gaps    0;

```

2351 TAGCTGATGCCGATAGTTAAGCAGCCCCGACACCCGCCAACCCGGTGACGGCCC 2410

Db 133 TGCTCTGAIGCCGCATGTTAAGCCAGCCCCGACACCCCGCCACACCCCGCTGACGGCCCC 192

2411 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470

Db 193 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGAGC 252

2471 TGCATGTGTGAGAGGTTTCAACCGTCATCACGAACGGCGAGACGAAGGGCCCTGTG 2530

Db 253 TCATGTGTACAGAGTTTCAACGTCATCACGAAACGCGCGAGACGAAAGGGCTTCGTG 312

2531 ATACGCCTATTTTATAGGTTAATGTCATGATAATAATGGTTCTTAG 2578

Db 313 ATACGCCTATTTTATAGGTTAAATGTCATGATATAATGTTCTTAG 360

Search completed: April 30, 2004, 15:28:51
Job time : 10412.6 secs

Job time : 10412.6 secs

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation; or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent lies
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 717 BP; 175 A; 165 C; 187 G; 190 T; 0 U; 0 Other;

XX Query Match 15.8%; Score 408.2; DB 7; Length 717;

XX Best Local Similarity 73.1%; Pred. No. 1.6e-113;

XX Matches 524; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 787 ATGATTTGCTACTATGATATGATTTGCTGCTTGGCCGGAATCTATCCAGTTTGA 846
 DB 1 ATGATTTGCTACTATGATATGATTTGCTGCTTGGCCGGAATCTATCCAGTTTGA 60
 QY 847 CTGCTGAAAGCGCTATCCAAATCGGCGAGATTTGCTGAAATGAGGCGATTTGCTGCGG 906
 DB 61 TTGGTCAAGCGAGCAATCCAAACCGCTGATGGCGGCTTTGGCGCAATCCCGGTGT 120
 QY 907 ATAGTGTATGACCAAGATATGACCGCTCATGCTGAGAAACCTGTGATGAGATTTGAT 966
 DB 121 TGCGTGTATGATGAGATCTCAGCGCAAGGTGGCGAAACCGCTACCAAAATTTGAT 180
 QY 967 GCAGCGAGAGAGATTTGCTGCGAAAGATCCATATCTATCTATTTTACCGCAT 1026
 DB 181 TCCGCGGCTGGCGCATCTGCCCGCAAAATTCATCTTACCCCACTTCACTGAT 240
 QY 1027 CCGCTGATGATGCAACCCCAACCGCATCCGATGCTGAGTGCATATTTCCAT 1086
 DB 241 CTTTGGTGAAGCATATCTTAAACCGATCCAAAGCGTGGAGCATCATTTCTCAG 300
 QY 1087 TAGCTTTATTAACCGGTGAGATTTCAAAACAAATGAATTTTGTATCGACCCGAG 1146
 DB 301 TAGCCCTGCTCATATTCACAGTCAACGCAATTTGAAGAGCTAGATGATTAAGAC 360
 QY 1147 CGCGTATATGTTTGTATCTTACTACTTACTCAAGAAAGAAATCGCGTGTCTTGTAT 1206
 DB 361 CGTGTATATGTTTGTACACAGAGATTTTCCAAAGTAAACAATTTGGGGTTTGTAT 420
 QY 1207 CATGTCGCGCAATTCAGCATGCGTGGCAAAATTAACGGGCTGCGGTTGAATACATCC 1266
 DB 421 CATGTGCGCAAAATTCAGCATGCAAAAACCTCGCGGACCGGTTAGTACATTTCT 480
 QY 1267 GTGGTACTATTTTCAATAGAAAAGTTTGTGAGCGAATTTATGATGCTTACGCG 1326
 DB 481 GTAGTGTCTTCTCCTTGAAGCTGGAGTTTGTGAGCGAATTTATGAGCTTACGCG 540
 QY 1327 AAAATTTGGCGTGAAGTAAATGCGATGACCATTAATACCGCGTTGATCAATACCGGGGTG 1386
 DB 541 AAAATTTGAGTGAAGTGAATGCGATGACCGTAAACGTCGGCTGATCATACCGGGGTGA 600

QY 1387 ATGCTGATGACCAATTAATATGCGCGATCTTGTAAAGCAAGACCAAGCCACC 1446
 DB 601 ATTGTGATGACCAATTAATATGCGCGCTTGTAAAGTAAAGCGCGCAAACTGCGAGC 660
 QY 1447 GTTATGCTGCGAGTCAATGAGTACCTATTTTCAACCGTTTACAGGTTAA 1503
 DB 661 GTTATGCTGCGAGTCAATGAGTACCTATTTTCAACCGTTTACAGGTTAA 717

RESULT 2

ACAS3529/c
 ID ACAS3529 standard; DNA; 1308 BP.

ACAS3529;

19-JUN-2003 (first entry)

Prokaryotic essential gene #35186.

Antisense; ds; prokaryotic essential gene; cell proliferation;

drug design; gene.

Vibrio cholerae.

MO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002MO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-0072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-1) ELITRA PHARM INC.

Mang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu RH;

WPI; 2003-029926/02.

P-PSDB; AB049559.

New antisense nucleic acids, useful for identifying proteins or screening

for homologous nucleic acids required for cellular proliferation to

isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 41399; 1766bp; English.

The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation; or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1308 BP; 319 A; 301 C; 324 G; 364 T; 0 U; 0 Other;

Query Match 9.7%; Score 249.2; DB 7; Length 1308;
 Best Local Similarity 70.6%; Pred. No. 8,2e-65;

Matches 332; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

DB 470 ACCCTCAGGTGTAATCAGCAGCAGCCTCAGCCCAACCTGACGGCTCGTACTTCAAGAT 411
 QY 23 ACTTGATGTTACTCGGCGGCACTTGCCCACTGCGCGGCTGACGTAATCTCAGAG 82
 DB 83 ATTAAGCCCATGACTAGCGTTTCATATAATGCTGCTCCGACGTAACCTTGAATGGCG 142
 DB 410 ACGATCGCCATCACTAAGGTTTCATACAAATCGTCAACGCACTGCCAGAACTCTG 351
 QY 143 ATACGCACTGCGCTTGGCCCTCTTGCTGAGGATCCGATTTCAATTTGCCATCGGGT 202
 DB 350 ATCCGAGTTGCTGTTTCCGCTTGTGACAAATCCCAATTTTACTTGTGCTTGAAGC 291
 QY 203 TGAATAATGAAATACGCTTAATACGCTTAATAAAGTACATTCAGTGAGTCTGCTGC 262
 DB 290 ACAAATGAAATTAACGCAACATTCGCAAGAGGCTGAGTGGGCGCATGCTGA 231
 QY 263 TCTAATAAATCAATGCTGCGCATCCGAAAGCGCAATGAGCACTGATTTCTTCTTG 322
 DB 230 GTCAATATATGATATCGGCATCCGAAAGGCAATGCGCAAGATGAGTATCTCTTGA 171
 QY 323 CGTACTTCTCTTAATAATCGCTAATGCTTATCATCTGCGCAATCAATTAATGAGCG 382
 DB 170 CGGATGGCATCAAGAGTCCGCTGCGATCCTCTTACGCGTACATCAATTCATGAGCG 111
 QY 383 ACCCTCAATCCGATACCAAGATGAAAGCGGCTGCATATTTGATTTTATAGGATCA 442
 DB 110 ACCGACATCTCGGATTAATAATGCTGTATCGCTTGCATCATTTGATTTTATAGCATCA 51
 QY 443 AGATCCAAATTAATCTCGGATTAATAAAGAGAGAGAAATAGCGCATCGCTCAT 492
 DB 50 AGATGAGCAGGCTTACGATGATGTGGGAGAGAAACAGGCGTGAATCAT 1

RESULT 3
 AB283655
 ID AB283655 standard; cDNA; 528 BP.

XX AB283655;
 AC 14-MAY-2003 (first entry)

XX Toxicologically relevant human nucleotide sequence #814.

XX Toxicologically relevant gene; toxicological response; gene; ss.

XX Homo sapiens.

XX MO2003016500-A2.

XX 27-FEB-2003.

XX 16-AUG-2002; 2002MO-US026514.

XX 16-AUG-2001; 2001US-0313080P.

XX (PHAS-1) PHASE-1 MOLECULAR TOXICOLOGY INC.

XX Nelt RE, Dunn RT, Adkins K, Pickett G, Kler LD, Schweiser K;

PI Allen P;
 XX WPI; 2003-268322/26.
 DR
 XX Determining a toxicological response to an agent, useful for screening of
 PT drugs, comprises comparing the expression profile of one or more human
 PT toxic response genes to a reference gene expression profile indicative of
 PT toxicity.

XX Claim 1; Page 255; 455pp; English.

XX The present invention describes a method (M1) for determining a
 CC toxicological response to an agent, which comprises comparing the
 CC expression profile of one or more human toxic response genes to a
 CC reference gene expression profile indicative of toxicity, and so
 CC determining the presence of a toxic response to the agent. Also
 CC described: (1) an array comprising one or more polynucleotides selected
 CC from the genes corresponding to the partial sequences given in AB282842
 CC to AB284764, or their fragments of at least 20 nucleotides, or homologues
 CC; and (2) determining if a gene putatively identified to be a toxic
 CC response gene plays a role on toxic response pathways by determining the
 CC expression profile of the gene after exposure of cells or a human subject
 CC to a known toxic pharmaceutical or industrial agent, comprising: (a)
 CC exposing cells to an agent or isolating cells from a human subject who
 CC was exposed to an agent; (b) obtaining the test gene expression profile
 CC for a putatively identified toxic response gene after exposure to a known
 CC toxic pharmaceutical or industrial agent; and (c) comparing the test
 CC profile to the expression profile of a gene with a similar function or
 CC comparing the test profile to the expression profile of that gene after
 CC exposure to other known toxic compounds. The methods are useful for
 CC predicting and determining toxicological responses on a cellular, organ
 CC or system level. The arrays comprising the human genes are useful for
 CC toxicological screening of drugs, pharmaceutical compounds and chemicals
 XX

SO Sequence 528 BP; 124 A; 139 C; 118 G; 144 T; 0 U; 3 Other;

Query Match 8.7%; Score 224.8; DB 7; Length 528;
 Best Local Similarity 99.1%; Pred. No. 1.4e-57;

Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTATGCGCCGATGTTAAGCCAGCCCGCAACCCCGCTGACGGGCGC 2410
 DB 227 TGCTCTGATGCGCCGATGTTAAGCCAGCCCGCAACCCCGCTGACGGGCGC 286
 QY 2411 TGAAGGCTTGTCTGCTCCGCGATCCGCTTACAGCAAGCTGTGACCGTCTCGGAGC 2470
 DB 287 TGAAGGCTTGTCTGCTCCGCGATCCGCTTACAGCAAGCTGTGACCGTCTCGGAGC 346
 QY 2471 TGATGTGTGACAGGTTTTCACCGTATCAACGAAACGGCGAGAGAAAGGCTCTG 2530
 DB 347 TGATGTGTGACAGGTTTTCACCGTATCAACGAAACGGCGAGAGAAAGGCTCTG 406
 QY 2531 ATAAGGCTATTTTATAGGTTAATGTCATGATATATGATGTTCTTCTAG 2578
 DB 407 ATAAGGCTATTTTATAGGTTAATGTCATGATATATGATGTTCTTCTAG 454

RESULT 4
 ABK42985
 ID ABK42985 standard; DNA; 878 BP.

XX ABK42985;

XX 21-MAY-2002 (first entry)

XX Genomic sequence #884 encoding novel human connective tissue polypeptide.

XX Human; connective tissue related disorder; cancer; gene therapy;

XX cytostatic; gene; ds.

XX Homo sapiens.

XX MO200155343-A1.

XX 02-AUG-2001.
 PD 17-JAN-2001; 2001WO-US001322.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-01806528P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216680P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218230P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236337P.

PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249247P.
 PR 17-NOV-2000; 2000US-0249248P.
 PR 17-NOV-2000; 2000US-0249249P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 06-DEC-2000; 2000US-0251656P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX

DR WPI; 2001-565190/63.
XX Nucleic acid encoding novel connective tissue associated polypeptides.
PT used in diagnosing, preventing, treating or ameliorating a disorder such
PT as cancer or rheumatoid arthritis.
PS Disclosure; SEQ ID NO 1872; 673bp; English.
XX
XX The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful in
CC gene therapy. ABK42102-ABK43116 represent genomic sequences encoding the
CC novel human connective tissue related polypeptides. Note: the sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;
Query Match 8.7%; Score 224.8; DB 4; Length 878;
Best Local Similarity 99.1%; Pred. No. 1.9e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TACGCTGATGCCGCTAGTTAAGCCAGCCCGACACCCCGCAACCCGCTGAGCGGCC 2410
DB 202 TCGCTGTATGCGCATAGTTAAGCCAGCCCGACACCCCGCAACCCGCTGAGCGGCC 261
QY 2411 TGACGGGCTGTGCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470
DB 262 TGACGGGCTGTGCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 321
QY 2471 TGCACTGTGTACAGGTTTTCACCGTCATCACCGCAACCGCGAGACAAAGGCGCTGTG 2530
DB 322 TGCACTGTGTACAGGTTTTCACCGTCATCACCGCAACCGCGAGACAAAGGCGCTGTG 381
QY 2531 ATAGCCCTATTTTATAGTTAATGTCATGATTAATGTTTCTTAG 2578
DB 382 ATAGCCCTATTTTATAGTTAATGTCATGATTAATGTTTCTTAG 429
RESULT 5
AA541856
ID AA541856 standard; DNA; 878 BP.
XX
XX AA541856;
AC
XX
DT 17-DEC-2001 (first entry)
XX
XX Genomic sequence #172 encoding novel human enzyme polypeptide.
DE
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KM ligase; hyperproliferative disorder; immunodeficiency disorder;
KM autoimmune disorder; neurological disorder; metabolic disorder;
KM inflammatory disorder; cardiovascular disorder; reproductive disorder;
KM blood-related disorder; infectious disorder; gene therapy; cytostatic;
XX anti arthritic; nephrotropic; anticoagulant; db.
XX
OS Homo sapiens.
XX
XX WO20015301-A2.
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001239.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186350P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR

PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 11-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-02321968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234233P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234999P.
PR 25-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR

PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250316P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-465566/50.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing, preventing,
 PT treating neural, immune system, muscular, reproductive, pulmonary,
 PT cardiovascular, renal, proliferative disorders and cancerous diseases.
 XX
 PS Disclosure; SEQ ID NO 1962; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences

CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerase or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
 CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
 CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
 CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
 CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
 CC infectious disorders (e.g. influenza). The polynucleotides of the
 CC invention can also be used in gene therapy. AA541685-AA542192 represent
 CC DNA sequences encoding for the novel human enzyme polypeptides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at fcp.wipo.int/publ/published_pot_sequences
 XX
 SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;

Query Match 8.7%; Score 224.8; DB 4; Length 878;
 Best Local Similarity 99.1%; Pred. No. 1.9e-57;
 Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCCCATAGTTTAAGCCAGCCCGGACACCCGCGCAACCCGCTGACGCGCC 2410
 Db 202 TGCTGTGATGCCCATAGTTTAAGCCAGCCCGGACACCCGCGCAACCCGCTGACGCGCC 261
 QY 2411 TGAAGGCTGTGTGCTCCCGGATCCGCTTACAGACAGCTGTGACCGCTCCGGAGC 2470
 Db 262 TGAAGGCTGTGTGCTCCCGGATCCGCTTACAGACAGCTGTGACCGCTCCGGAGC 321
 QY 2471 TGCATGTGTGAGAGTTTTCACCGTCATCACGAAAGCGCGAGAGAAAGGCTCTGTG 2530
 Db 322 TGCATGTGTGAGAGTTTTCACCGTCATCACGAAAGCGCGAGAGAAAGGCTCTGTG 381
 QY 2531 ATACGCTTATTTTATATAGTTTATGTGATGATATATATGTTTCTTATG 2578
 Db 382 ATACGCTTATTTTATATAGTTTATGTGATGATATATATGTTTCTTATG 429

RESULT 6
 AAK85435
 ID AAK85435 standard; DNA; 878 BP.
 XX
 AC AAK85435;
 XX
 DE 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40247.
 XX
 DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 XX
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001354.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214986P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880F.
PR 11-JUL-2000; 2000US-02117487P.
PR 11-JUL-2000; 2000US-02117496F.
PR 14-JUL-2000; 2000US-0218190F.
PR 26-JUL-2000; 2000US-0220964F.
PR 26-JUL-2000; 2000US-0220964F.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266F.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-022679P.
PR 22-AUG-2000; 2000US-0226811P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0233397P.
PR 14-SEP-2000; 2000US-0233398P.
PR 14-SEP-2000; 2000US-0233399P.
PR 14-SEP-2000; 2000US-0233400P.
PR 14-SEP-2000; 2000US-0233401P.
PR 14-SEP-2000; 2000US-023363P.
PR 14-SEP-2000; 2000US-0233654P.
PR 14-SEP-2000; 2000US-0233655P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-023497P.
PR 25-SEP-2000; 2000US-023498P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0235327P.
PR 29-SEP-2000; 2000US-0235367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250300P.
PR 01-DEC-2000; 2000US-0250301P.
PR 01-DEC-2000; 2000US-0250310P.
PR 05-DEC-2000; 2000US-0251039P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251909P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI, 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 40247; 3071pp + Sequence Listing; English.
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK91521. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the

| | | | |
|---------------------------|--|--------------|------------------|
| | PR | 08-NOV-2000; | 2000US-0246524P- |
| | PR | 08-NOV-2000; | 2000US-0246525P. |
| | PR | 08-NOV-2000; | 2000US-0246525P. |
| | PR | 08-NOV-2000; | 2000US-0246527P. |
| | PR | 08-NOV-2000; | 2000US-0246528P. |
| | PR | 08-NOV-2000; | 2000US-0246532P. |
| | PR | 08-NOV-2000; | 2000US-0246609P. |
| | PR | 08-NOV-2000; | 2000US-0246610P. |
| | PR | 08-NOV-2000; | 2000US-0246613P. |
| | PR | 17-NOV-2000; | 2000US-0246207P. |
| | PR | 17-NOV-2000; | 2000US-0249208P. |
| | PR | 17-NOV-2000; | 2000US-0249209P. |
| | PR | 17-NOV-2000; | 2000US-0249210P. |
| | PR | 17-NOV-2000; | 2000US-0249211P. |
| | PR | 17-NOV-2000; | 2000US-0249212P. |
| | PR | 17-NOV-2000; | 2000US-0249213P. |
| | PR | 17-NOV-2000; | 2000US-0249214P. |
| | PR | 17-NOV-2000; | 2000US-0249215P. |
| | PR | 17-NOV-2000; | 2000US-0249216P. |
| | PR | 17-NOV-2000; | 2000US-0249217P. |
| | PR | 17-NOV-2000; | 2000US-0249300P. |
| | PR | 01-DEC-2000; | 2000US-0250160P. |
| | PR | 01-DEC-2000; | 2000US-0250391P. |
| | PR | 05-DEC-2000; | 2000US-0251030P. |
| | PR | 05-DEC-2000; | 2000US-0251988P. |
| | PR | 05-DEC-2000; | 2000US-0254792P. |
| | PR | 06-DEC-2000; | 2000US-0251856P. |
| | PR | 08-DEC-2000; | 2000US-0251868P. |
| | PR | 08-DEC-2000; | 2000US-0251869P. |
| | PR | 08-DEC-2000; | 2000US-0251989P. |
| | PR | 11-DEC-2000; | 2000US-0254097P. |
| | PR | 05-JAN-2001; | 2001US-0259678P. |
| PX | (HUMA-) HUMAN GENOME SCT INC. | | |
| PY | Rosen CA, Barash SC, Ruben SM, | | |
| DZ | WPI, 2001-465570/50. | | |
| XX | | | |
| PX | | | |
| PY | Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition. | | |
| DZ | | | |
| XX | | | |
| PX | | | |
| PY | Disclosure; SEQ ID NO 10029; 129Tbp + Sequence Listing; English. | | |
| DZ | | | |
| XX | The present invention provides the protein and coding sequences of a number of human reproductive system related antigens These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention | | |
| CX | | | |
| CC | | | |
| CC | | | |
| CC | | | |
| CC | | | |
| SC | Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other; | | |
| SQ | | | |
| Query Match: | 8.7% Score 224.8; DB 4; Length 878; | | |
| Best Local Similarity | 99.1%; Pred.No.19e-57; | | |
| Matches 226; Conservative | 0; Mismatches 2; Indels 0; Gaps 0; | | |
| OY | 2351 TAGCGTGATGCGCATAGTTAAACCAAGCACCACGCCAACAACCCGTGAACGGCCC 2410U | | |
| Db | 202 TGCTGTATTGCAGTCATTAAGTAGCACAGCCCCGACCAACCCGCGTGAACGGCCC 261 | | |
| OY | 2411 TGACGGCGCTTGTCTGCTCCCGGCATCCGCTTACAGAACAAGCTGTGAACCGCTCTCCGGGAGC 2470U | | |
| Db | 262 TGACCGGGCTTGTCTGCTCCCGGCATCCGCTTACAGAACAAGCTGTGAACCGCTCTCCGGGAGC 321 | | |

| | | | | |
|----------|---|------|---|------|
| Oy | | 2471 | TGATGTGTGCAGAGTTTTCACCGTGATCACCCGAAGCGGAGACGAAGAAGGGCTCGTG | 2539 |
| Dd | | 322 | TGATGTGTGTGCAGAGGTTTCAACCGTCATCACCGGAAGCGGAGAACGAAGGGCTCGTG | 381 |
| Oy | | 2531 | ATAACGCTATTATTAAGTTAATGTCATGATGATTAATAATGTTTCTTAG | 2578 |
| Dd | | 382 | ATACGCCATTATTTATATGATTAATGATCATGATATATAATAGTTTCTTAG | 429 |
| RESULT 8 | | | | |
| ID | AL07345 | | | |
| XX | AL07345 standard; DNA; 878 BP. | | | |
| AC | AA07345; | | | |
| XX | | | | |
| DT | 21-NOV-2001 (first entry) | | | |
| XX | | | | |
| DE | Human reproductive system related antigen DNA SEQ ID NO: 10033. | | | |
| XX | | | | |
| KW | Human; reproductive system related antigen; reproductive system disorder; | | | |
| XV | cancer; Gene therapy; ds. | | | |
| XX | | | | |
| OS | Homo sapiens. | | | |
| XX | | | | |
| PN | WO200155320-A2. | | | |
| XX | | | | |
| PD | 02-AUG-2001. | | | |
| XX | | | | |
| PF | 17-JAN-2001; 2001WO-US001339. | | | |
| XX | | | | |
| PR | 31-JAN-2000; 2000US-0179065P. | | | |
| PR | 04-FEB-2000; 2000US-0180628P. | | | |
| PR | 24-FEB-2000; 2000US-0184664P. | | | |
| PR | 02-MAR-2000; 2000US-0186350P. | | | |
| PR | 16-MAR-2000; 2000US-0189874P. | | | |
| PR | 17-MAR-2000; 2000US-0190076P. | | | |
| PR | 18-APR-2000; 2000US-0198123P. | | | |
| PR | 19-MAY-2000; 2000US-0205515P. | | | |
| PR | 07-JUN-2000; 2000US-0209467P. | | | |
| PR | 28-JUN-2000; 2000US-0214886P. | | | |
| PR | 30-JUN-2000; 2000US-0215135P. | | | |
| PR | 07-JUL-2000; 2000US-0216647P. | | | |
| PR | 07-JUL-2000; 2000US-0216880P. | | | |
| PR | 11-JUL-2000; 2000US-0217487P. | | | |
| PR | 14-JUL-2000; 2000US-0218290P. | | | |
| PR | 26-JUL-2000; 2000US-0220963P. | | | |
| PR | 26-JUL-2000; 2000US-0220964P. | | | |
| PR | 14-AUG-2000; 2000US-0224518P. | | | |
| PR | 14-AUG-2000; 2000US-0224519P. | | | |
| PR | 14-AUG-2000; 2000US-0225213P. | | | |
| PR | 14-AUG-2000; 2000US-0225214P. | | | |
| PR | 14-AUG-2000; 2000US-0225266P. | | | |
| PR | 14-AUG-2000; 2000US-0225267P. | | | |
| PR | 14-AUG-2000; 2000US-0225268P. | | | |
| PR | 14-AUG-2000; 2000US-0225270P. | | | |
| PR | 14-AUG-2000; 2000US-0225447P. | | | |
| PR | 14-AUG-2000; 2000US-0225757P. | | | |
| PR | 14-AUG-2000; 2000US-0225758P. | | | |
| PR | 14-AUG-2000; 2000US-0225759P. | | | |
| PR | 18-AUG-2000; 2000US-0226279P. | | | |
| PR | 22-AUG-2000; 2000US-0226681P. | | | |
| PR | 22-AUG-2000; 2000US-0226686P. | | | |
| PR | 22-AUG-2000; 2000US-0227182P. | | | |
| PR | 23-AUG-2000; 2000US-0227009P. | | | |
| PR | 30-AUG-2000; 2000US-0228924P. | | | |
| PR | 01-SEP-2000; 2000US-0229287P. | | | |
| PR | 01-SEP-2000; 2000US-0229343P. | | | |
| PR | 01-SEP-2000; 2000US-0229344P. | | | |
| PR | 01-SEP-2000; 2000US-0229345P. | | | |
| PR | 05-SEP-2000; 2000US-0229509P. | | | |
| PR | 05-SEP-2000; | | | |

[illegible]

```

PR      17-NOV-2000; 2000US-0249218P.
PR      17-NOV-2000; 2000US-0249224P.
PR      17-NOV-2000; 2000US-0249245P.
PR      17-NOV-2000; 2000US-0249264P.
PR      17-NOV-2000; 2000US-0249265P.
PR      17-NOV-2000; 2000US-0249297P.
PR      17-NOV-2000; 2000US-0249299P.
PR      17-NOV-2000; 2000US-0249300P.
PR      01-DEC-2000; 2000US-0250160P.
PR      01-DEC-2000; 2000US-0250381P.
PR      05-DEC-2000; 2000US-0251030P.
PR      05-DEC-2000; 2000US-0251988P.
PR      05-DEC-2000; 2000US-0256719P.
PR      06-DEC-2000; 2000US-0251479P.
PR      08-DEC-2000; 2000US-0251856P.
PR      08-DEC-2000; 2000US-0251858P.
PR      08-DEC-2000; 2000US-0251869P.
PR      08-DEC-2000; 2000US-0251989P.
PR      11-DEC-2000; 2000US-0254097P.
PR      05-JAN-2001; 2001US-0259678P.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Rosen CA, Barash SC, Ruben SM;
XX
XX      WPI; 2001-465570/50.
XX
XX      Isolated nucleic acid molecule encoding a reproductive system antigen is
XX      used in preventing, treating or ameliorating a medical condition.
XX
XX      Disclosure; SEQ ID NO 10033; 1297bp + Sequence Listing; English.
XX
XX      The present invention provides the protein and coding sequences of a
XX      number of human reproductive system related antigens. These can be used
XX      in the prevention and treatment of reproductive system disorders,
XX      including cancer. The present sequence is a genomic sequence encoding a
XX      protein of the invention
CC
CC      Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;
SQ
Query Match          8.7%; Score 224.8; DB 4; Length 878;
Best Local Similarity 99.1%; Pred. No. 1.9e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      2351 TAGCGTGAAGCCGCAATTAAGCAGACCCCGCACACCCTGCCAACAACCGCTGACGGCCC 2410
DB      202 TGCTGTATGCCGCATGTTAAAGCAGACCCCGCAACAACCGCTGACGGCCCC 261
QY      2411 TGAAGGGCTTCCTGCTCCCGGCAATCCGCTTAAGAAGTGACCGTCCCGGAGC 2470
DB      262 TGAAGGGGTTGCTGCTCCCGGCAATCCGCTTAAGAAGTGACCGTCCCGGAGC 321
QY      2471 TGCATGTGTCAAGAGTTTTCACCGTCAACCGAAACGCGGAGACGAAGAGGCTCGTG 2530
DB      322 TGCAATGTGTCAAGAGTTTTCACCGTCAACCGAAACGCGGAGACGAAGAGGCTCGTG 381
QY      2531 ATAGCGCTATTATTTATAGCTTAATGCTAATGATAATATGTTCTTTG 2578
DB      382 ATAGCGCTATTATTTATAGCTTAATGCTAATGATAATATGTTCTTTG 429
RESULT 9
AAS34682
ID      AAS34682 standard; DNA; 878 BP.
AC      AAS34682;
XX
XX      17-DEC-2001 (first entry)
XX
XX      Human DNA for a novel foetal antigen, SEQ ID No 2106.
XX
XX      Human; foetal tissue antigen; ds; antiinflammatory; neuroprotective;
KW
```

| | | | |
|----|---|----|-------------------------------|
| KW | immunomodulator; cardiovascular; cytosolic; nephrothropic; | PR | 14-SEP-2000; 2000US-0233063P. |
| KW | cardiovascular; autoimmune disease; rheumatoid arthritis; | PR | 14-SEP-2000; 2000US-0233064P. |
| KW | hyperproliferative disorder; breast neoplasm; cancer; | PR | 14-SEP-2000; 2000US-0233065P. |
| KW | cardiovascular disorder; cardiac arrest; cerebrovascular disorder; | PR | 21-SEP-2000; 2000US-0234223P. |
| KW | cerebral ischemia; angioneurosis; nervous system disorder; | PR | 21-SEP-2000; 2000US-0234274P. |
| KW | Alzheimer's disease; infection; ocular disorder; corneal infection; | PR | 25-SEP-2000; 2000US-0234997P. |
| KW | wound healing; epithelial cell proliferation; food additive. | PR | 25-SEP-2000; 2000US-0234998P. |
| OS | Homo sapiens. | PR | 26-SEP-2000; 2000US-0235484P. |
| PN | WO200155312-A2. | PR | 27-SEP-2000; 2000US-0235834P. |
| XX | | PR | 27-SEP-2000; 2000US-0235836P. |
| XX | | PR | 29-SEP-2000; 2000US-0236327P. |
| PD | 02-AUG-2001. | PR | 29-SEP-2000; 2000US-0236367P. |
| XX | | PR | 29-SEP-2000; 2000US-0236368P. |
| XX | | PR | 29-SEP-2000; 2000US-0236369P. |
| PF | 17-JAN-2001; 2001WO-US001321. | PR | 29-SEP-2000; 2000US-0236370P. |
| XX | | PR | 02-OCT-2000; 2000US-0236802P. |
| XX | | PR | 02-OCT-2000; 2000US-0237037P. |
| PR | 31-JAN-2000; 2000US-0179065P. | PR | 02-OCT-2000; 2000US-0237038P. |
| PR | 04-FEB-2000; 2000US-0180628P. | PR | 02-OCT-2000; 2000US-0237039P. |
| PR | 24-FEB-2000; 2000US-0184664P. | PR | 02-OCT-2000; 2000US-0237040P. |
| PR | 02-MAR-2000; 2000US-0186350P. | PR | 13-OCT-2000; 2000US-0239337P. |
| PR | 16-MAR-2000; 2000US-0188874P. | PR | 13-OCT-2000; 2000US-0239337P. |
| PR | 17-MAR-2000; 2000US-0190076P. | PR | 20-OCT-2000; 2000US-0240960P. |
| PR | 18-APR-2000; 2000US-0198123P. | PR | 20-OCT-2000; 2000US-0241212P. |
| PR | 19-MAY-2000; 2000US-0205155P. | PR | 20-OCT-2000; 2000US-0241785P. |
| PR | 07-JUN-2000; 2000US-0209467P. | PR | 20-OCT-2000; 2000US-0241787P. |
| PR | 28-JUN-2000; 2000US-0214886P. | PR | 20-OCT-2000; 2000US-0241808P. |
| PR | 30-JUN-2000; 2000US-0215135P. | PR | 20-OCT-2000; 2000US-0241809P. |
| PR | 07-JUL-2000; 2000US-0216647P. | PR | 20-OCT-2000; 2000US-0241826P. |
| PR | 07-JUL-2000; 2000US-0216880P. | PR | 01-NOV-2000; 2000US-0244617P. |
| PR | 11-JUL-2000; 2000US-0217487P. | PR | 08-NOV-2000; 2000US-0246474P. |
| PR | 14-JUL-2000; 2000US-0218290P. | PR | 08-NOV-2000; 2000US-0246475P. |
| PR | 26-JUL-2000; 2000US-0220963P. | PR | 08-NOV-2000; 2000US-0246476P. |
| PR | 26-JUL-2000; 2000US-0220964P. | PR | 08-NOV-2000; 2000US-0246477P. |
| PR | 14-AUG-2000; 2000US-0224518P. | PR | 08-NOV-2000; 2000US-0246478P. |
| PR | 14-AUG-2000; 2000US-0224519P. | PR | 08-NOV-2000; 2000US-0246523P. |
| PR | 14-AUG-2000; 2000US-0225213P. | PR | 08-NOV-2000; 2000US-0246524P. |
| PR | 14-AUG-2000; 2000US-0225214P. | PR | 08-NOV-2000; 2000US-0246525P. |
| PR | 14-AUG-2000; 2000US-0225266P. | PR | 08-NOV-2000; 2000US-0246526P. |
| PR | 14-AUG-2000; 2000US-0225267P. | PR | 08-NOV-2000; 2000US-0246527P. |
| PR | 14-AUG-2000; 2000US-0225268P. | PR | 08-NOV-2000; 2000US-0246528P. |
| PR | 14-AUG-2000; 2000US-0225270P. | PR | 08-NOV-2000; 2000US-0246532P. |
| PR | 14-AUG-2000; 2000US-0225447P. | PR | 08-NOV-2000; 2000US-0246609P. |
| PR | 14-AUG-2000; 2000US-0225757P. | PR | 08-NOV-2000; 2000US-0246610P. |
| PR | 14-AUG-2000; 2000US-0225758P. | PR | 08-NOV-2000; 2000US-0246611P. |
| PR | 14-AUG-2000; 2000US-0225759P. | PR | 08-NOV-2000; 2000US-0246613P. |
| PR | 18-AUG-2000; 2000US-0226279P. | PR | 17-NOV-2000; 2000US-0249207P. |
| PR | 22-AUG-2000; 2000US-0226681P. | PR | 17-NOV-2000; 2000US-0249208P. |
| PR | 22-AUG-2000; 2000US-0226688P. | PR | 17-NOV-2000; 2000US-0249209P. |
| PR | 22-AUG-2000; 2000US-0227182P. | PR | 17-NOV-2000; 2000US-0249210P. |
| PR | 23-AUG-2000; 2000US-0227182P. | PR | 17-NOV-2000; 2000US-0249211P. |
| PR | 30-AUG-2000; 2000US-0228924P. | PR | 17-NOV-2000; 2000US-0249212P. |
| PR | 01-SEP-2000; 2000US-0229287P. | PR | 17-NOV-2000; 2000US-0249213P. |
| PR | 01-SEP-2000; 2000US-0229343P. | PR | 17-NOV-2000; 2000US-0249214P. |
| PR | 01-SEP-2000; 2000US-0229344P. | PR | 17-NOV-2000; 2000US-0249215P. |
| PR | 01-SEP-2000; 2000US-0229345P. | PR | 17-NOV-2000; 2000US-0249216P. |
| PR | 05-SEP-2000; 2000US-0229509P. | PR | 17-NOV-2000; 2000US-0249217P. |
| PR | 05-SEP-2000; 2000US-0229513P. | PR | 17-NOV-2000; 2000US-0249218P. |
| PR | 06-SEP-2000; 2000US-0230437P. | PR | 17-NOV-2000; 2000US-0249244P. |
| PR | 06-SEP-2000; 2000US-0230438P. | PR | 17-NOV-2000; 2000US-0249245P. |
| PR | 08-SEP-2000; 2000US-0231242P. | PR | 17-NOV-2000; 2000US-0249246P. |
| PR | 08-SEP-2000; 2000US-0231243P. | PR | 17-NOV-2000; 2000US-0249247P. |
| PR | 08-SEP-2000; 2000US-0231244P. | PR | 17-NOV-2000; 2000US-0249248P. |
| PR | 08-SEP-2000; 2000US-0231249P. | PR | 17-NOV-2000; 2000US-0249249P. |
| PR | 08-SEP-2000; 2000US-0231413P. | PR | 17-NOV-2000; 2000US-0249297P. |
| PR | 08-SEP-2000; 2000US-0231414P. | PR | 17-NOV-2000; 2000US-0249299P. |
| PR | 08-SEP-2000; 2000US-0230801P. | PR | 01-DEC-2000; 2000US-0250300P. |
| PR | 08-SEP-2000; 2000US-0230801P. | PR | 01-DEC-2000; 2000US-0250301P. |
| PR | 12-SEP-2000; 2000US-0231968P. | PR | 05-DEC-2000; 2000US-0251030P. |
| PR | 14-SEP-2000; 2000US-0232398P. | PR | 05-DEC-2000; 2000US-0251030P. |
| PR | 14-SEP-2000; 2000US-0232398P. | PR | 05-DEC-2000; 2000US-0251030P. |
| PR | 14-SEP-2000; 2000US-0232399P. | PR | 05-DEC-2000; 2000US-0251030P. |
| PR | 14-SEP-2000; 2000US-0232400P. | PR | 05-DEC-2000; 2000US-0251030P. |
| PR | 14-SEP-2000; 2000US-0232401P. | PR | 08-DEC-2000; 2000US-0251856P. |

Query Match 8.7%; Score 224.8; DB 7; Length 878;
 Best Local Similarity 99.1%; Pred. No. 1.9e-57;
 Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 2351 TACGCTGATGCGCATAGTTAAGCCAGCCGACCCGACCAACCCGCTGACGCGCC 2410
 202 TGCTCTGATGCGCATAGTTAAGCCAGCCGACCCGACCAACCCGCTGACGCGCC 261

QY 2411 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCTCCGGAGC 2470
 262 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCTCCGGAGC 321

DB 2471 TGACATGTGTACAGAGTTTTCACCGCTATCACCGAAGCGCGAGACGAAAGGCGCTG 2530
 322 TGACATGTGTACAGAGTTTTCACCGCTATCACCGAAGCGCGAGACGAAAGGCGCTG 381

QY 2531 ATAGCCCTATTTTATAGTTAATGTCAATGATATATATGTTTCTTAG 2578
 382 ATAGCCCTATTTTATAGTTAATGTCAATGATATATATGTTTCTTAG 429

DB

RESULT 11

ADA41575 standard; DNA; 878 BP.

XX ADA41575;
 AC 20-NOV-2003 (first entry)
 XX Human secreted protein related DNA.
 DE Human; secreted protein; cancer; hyperproliferative disorder;
 XX Rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
 XX anaemia; allergic reaction; asthma; cardiovascular disorder;
 XX wound healing; cytostatic; immunosuppressive; nocrotropic; neuroprotective;
 XX antiviral; anti-allergic; hepatocytropic; antidiabetic; antiinflammatory;
 XX vulnery; cardiant; gene therapy; ds.
 XX Homo sapiens.
 OS
 XX WO2002102993-A2.
 PN 27-DEC-2002.
 PD 19-MAR-2002; 2002WO-US008123.
 PF 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-175238/17.
 DR New human secreted proteins and nucleic acid molecules, useful for
 XX preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating cancer or other hyperproliferative disorder,
 PT asthma, allergies or AIDS.
 XX
 PS Disclosure, SEQ ID NO 1958; 3205bp; English.
 XX
 XX The invention relates to novel genes ADA39629-ADA40565 and proteins
 CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
 CC fragments, and agonists or antagonists that bind to the polypeptide are
 CC useful for preparing a diagnostic or pharmaceutical composition for
 CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune

CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. The sequences given in
 CC ADA40803-ADA4165 represent DNA sequences related to human secreted
 CC proteins. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX

SO Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;

Query Match 8.7%; Score 224.8; DB 7; Length 878;
 Best Local Similarity 99.1%; Pred. No. 1.9e-57;
 Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCATAGTTAAGCCAGCCGACCCGACCAACCCGCTGACGCGCC 2410
 DB 202 TGCTCTGATGCGCATAGTTAAGCCAGCCGACCCGACCAACCCGCTGACGCGCC 261

QY 2411 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCTCCGGAGC 2470
 DB 262 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCTCCGGAGC 321

QY 2471 TGACATGTGTACAGAGTTTTCACCGCTATCACCGAAGCGCGAGACGAAAGGCGCTG 2530
 DB 322 TGACATGTGTACAGAGTTTTCACCGCTATCACCGAAGCGCGAGACGAAAGGCGCTG 381

QY 2531 ATAGCCCTATTTTATAGTTAATGTCAATGATATATATGTTTCTTAG 2578
 DB 382 ATAGCCCTATTTTATAGTTAATGTCAATGATATATATGTTTCTTAG 429

RESULT 12

ACC50906 standard; cDNA; 878 BP.

XX ACC50906;
 AC 12-JUN-2003 (first entry)
 XX Human secreted protein BAC clone SEQ ID NO 1086.
 DE Cardiant; antiarrhythmic; antiarteriosclerotic; vasotrophic; cytostatic;
 XX Cardiant; antiinflammatory; nocrotropic; neuroprotective;
 XX antiparkinsonian; gene therapy; human; cardiovascular disorder; ss.
 XX Homo sapiens.
 OS
 XX WO200295010-A2.
 PN 28-NOV-2002.
 PD 19-MAR-2002; 2002WO-US009785.
 PF 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX

DR WPI; 2003-129429/12.
XX Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX
PS Disclosure; SEQ ID NO 1086; 1881bp; English.
XX
XX The present invention relates to novel human secreted proteins (ABR47633-
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
CC and their coding sequences are useful for the preparation of a diagnostic
CC or pharmaceutical composition for diagnosing or treating a cardiovascular
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
CC arteriosclerosis and myocardial ischemia), neural disorders, coronary
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, renal disorders,
CC proliferative disorders and/or cancerous diseases and conditions, for
CC wound healing and epithelial cell proliferation, to treat inflammation or
CC infection, for treating thrombosis and arteriosclerosis, for treating or
CC preventing neural damage which occurs in neuronal disorders or
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, to enhance bone and periodontal regeneration and aid in tissue
CC transplants or bone grafts, to prevent skin aging or hair loss, to
CC stimulate growth and differentiation of hematopoietic cells and bone
CC marrow cells when used in combination with other cytokines, to maintain
CC organs before transplantation or for supporting cell culture of primary
CC tissues, to increase or decrease differentiation or proliferation of
CC embryonic stem cells, or to modulate mammalian characteristics or
CC metabolism. The present sequence was used to illustrate the invention.
CC Note: The sequence data for this patent was published in electronic
CC format and is available from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;
Query Match 8.7%; Score 224.8; DB 7; Length 878;
Best Local Similarity 99.1%; Pred. No. 1.9e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TACGCTGATGCGCATGATTAAAGCCAGCCCGCAACCCCGCTGACGGCCC 2410
DB 202 TGCCTGTATGCCCATAGTTAAAGCCAGCCCGCAACCCCGCTGACGGCCC 261
QY 2411 TGACGGGCTGTGCTGCTCCCGCATCCGCTTACAGACAAAGCTGTGACCGTCCGGGAGC 2470
DB 262 TGACGGGCTGTGCTGCTCCCGCATCCGCTTACAGACAAAGCTGTGACCGTCCGGGAGC 321
QY 2471 TGCAATGTGTCAGAGGTTTTCACCGTCATCAACCGAAACGGCGAGAAAGGCGCTCGTG 2530
DB 322 TGCAATGTGTCAGAGGTTTTCACCGTCATCAACCGAAACGGCGAGAAAGGCGCTCGTG 381
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578
DB 382 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 429
RESULT 13
ABZ71509
ID ABZ71509 standard; DNA, 878 BP.
XX
AC ABZ71509;
XX
DT 03-APR-2003 (first entry)
XX
DE Secreted protein gene 74 genomic fragment HMOG30, SEQ ID NO:619.
XX
XX Human; secreted protein; digestive disorder; gastrointestinal disorder;
XX mouth; oesophagus; stomach; small intestine; large intestine; liver;
XX biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;
XX immune disorder; inflammation; infection; wound healing; drug screening;
XX chromosome identification; chromosome mapping; cytostatic; gene therapy;
XX antiinflammatory; immunosuppressive; valvulopathy; chromosome 15; gene; ds.

OS Homo sapiens.
XX
XX WO200276488-A1.
XX
XX 03-OCT-2002.
XX
XX 19-MAR-2002; 2002WC-US008276.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX 19-JUL-2001; 2001US-0306171P.
XX 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-029900/02.
XX
XX New human secreted proteins and nucleic acids, useful for detecting,
PT preventing, diagnosing, prognosticating, treating and/or ameliorating
PT e.g. gastrointestinal diseases and disorders, or cancers.
XX
XX Disclosure; Page 1184; 1216bp; English.
XX
XX ABR71190-ABZ71478 represent cDNAs corresponding to 178 human secreted
CC protein genes, and ABR00011-ABP00299 represent the proteins they encode.
CC ABR71479-ABZ71540 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening, and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing, treating, ameliorating or preventing
CC digestive disorders. Such conditions include disorders of the mouth,
CC oesophagus, stomach, small intestine, large intestine, liver, biliary
CC tract and pancreas, and include cancers of these organs and tissues. The
CC secreted proteins and their nucleic acids may also be used in the
CC treatment of immune disorders, inflammation, infection,
CC hyperproliferative disorders, and to promote wound healing. Nucleic acids
CC of the invention may be used for chromosome identification, chromosome
CC mapping, in gene therapy, for identifying individuals from minute
CC biological samples, as hybridisation probes, and as molecular weight
CC markers. The present sequence represents a human secreted protein genomic
CC fragment referred to in the disclosure of the invention
XX
SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;
Query Match 8.7%; Score 224.8; DB 7; Length 878;
Best Local Similarity 99.1%; Pred. No. 1.9e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TACGCTGATGCGCATGATTAAAGCCAGCCCGCAACCCCGCTGACGGCCC 2410
DB 202 TGCCTGTATGCCCATAGTTAAAGCCAGCCCGCAACCCCGCTGACGGCCC 261
QY 2411 TGACGGGCTGTGCTGCTCCCGCATCCGCTTACAGACAAAGCTGTGACCGTCCGGGAGC 2470
DB 262 TGACGGGCTGTGCTGCTCCCGCATCCGCTTACAGACAAAGCTGTGACCGTCCGGGAGC 321
QY 2471 TGCAATGTGTCAGAGGTTTTCACCGTCATCAACCGAAACGGCGAGAAAGGCGCTCGTG 2530
DB 322 TGCAATGTGTCAGAGGTTTTCACCGTCATCAACCGAAACGGCGAGAAAGGCGCTCGTG 381
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578
DB 382 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 429
RESULT 14
ADB91870
ID ADB91870 standard; DNA, 878 BP.
XX
AC ADB91870;

XX 04-DEC-2003 (first entry)
DT Human secreted protein related DNA #SEQ ID 816.
XX
XX Secreted protein; gene therapy; antidiabetic; diabetes; human; ds.
XX
XX Homo sapiens.
XX
XX WO2003004622-A2.
XX
XX 16-JAN-2003.
XX
XX 19-MAR-2002; 2002WO-US008124.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX
XX 19-JUL-2001; 2001US-0306171P.
XX
XX 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-229407/22.
XX
XX Nucleic acid encoding a human secreted protein is useful in diagnosing or
PT treating diabetes or conditions related to diabetes.
XX
XX Disclosure; SEQ ID NO 816; 1537bp; English.
XX
XX The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;
SQ
Query Match 8.7%; Score 224.8; DB 8; Length 878;
Best Local Similarity 99.1%; Pred. No. 1.9e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TACGCTGATGCGCATGTTAAGCAGAGCCCGGACACCCGCAACACCGCTGAGCGGCC 2410
DB 202 TGCTTGATGCGCATGTTAAGCAGAGCCCGGACACCCGCAACACCGCTGAGCGGCC 261
QY 2411 TGACGCGGCTTGTCTCTCCCGCATCGCTTACAGACAGAGTGTGACCGTCTCCGGAGC 2470
DB 262 TGACGCGGCTTGTCTCTCCCGCATCGCTTACAGACAGAGTGTGACCGTCTCCGGAGC 321
QY 2471 TGACGCTGTCAGAGGTTTTCACCGTCATCACGAAACCGGAGAGCGAAAGGCGCTG 2510
DB 322 TGACGCTGTCAGAGGTTTTCACCGTCATCACGAAACCGGAGAGCGAAAGGCGCTG 381
QY 2531 ATACGCTATTTTATAGATTATGTCATGATTAATATGTTCTTAA 2578
DB 382 ATACGCTATTTTATAGATTATGTCATGATTAATATGTTCTTAA 429

RESULT 15
ADB61141
ID ADB61141 standard; DNA; 878 BP.

XX ADB61141;
AC
XX 04-DEC-2003 (first entry)
DT
XX
XX Connective tissue related genomic DNA #884.
XX
XX Cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;
XX antiarteriosclerotic; immunosuppressive; antirheumatic; antiarthritis;
XX antiinflammatory; antiallergic; antiaesthetic; dermatological;
XX nephrotoxic; virucide; fungicide; antibacterial; antiparasitic;
XX gene therapy; ds; connective tissues disorder; rheumatoid arthritis;
XX systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
XX cancer metastasis; neoplasia; leukaemia; neurodegenerative disorder;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disease;
XX atherosclerosis; myocarditis; cardiopulmonary bypass complication;
XX autoimmune disease; multiple sclerosis; allergic reaction; asthma;
XX rhinitis; eczema; inflammatory condition; Crohn's disease;
XX gastrointestinal disorder; inflammatory bowel disease;
XX organ transplant rejection; immune system disorder; Bruton's disease;
XX X-linked lymphoproliferative syndrome;
XX B-cell lymphoproliferative disorder; HIV; AIDS; infection;
XX chromosome identification; chromosome mapping;
XX connective tissue related polynucleotide; gene; ds.
XX
XX Homo sapiens.
XX
XX US2003054375-A1.
XX
XX 20-MAR-2003.
XX
XX 07-MAR-2002; 2002US-00092154.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 02-MAR-2000; 2000US-0184664P.
XX
XX 16-MAR-2000; 2000US-0186350P.
XX
XX 17-MAR-2000; 2000US-0189874P.
XX
XX 18-APR-2000; 2000US-0190076P.
XX
XX 19-MAY-2000; 2000US-0198123P.
XX
XX 07-JUN-2000; 2000US-0205151P.
XX
XX 28-JUN-2000; 2000US-0209467P.
XX
XX 30-JUN-2000; 2000US-0214886P.
XX
XX 07-JUL-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 11-JUL-2000; 2000US-0217496P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220863P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225271P.
XX
XX 14-AUG-2000; 2000US-0225275P.
XX
XX 14-AUG-2000; 2000US-0225278P.
XX
XX 14-AUG-2000; 2000US-0225279P.
XX
XX 14-AUG-2000; 2000US-0226279P.
XX
XX 14-AUG-2000; 2000US-0226281P.
XX
XX 22-AUG-2000; 2000US-0226688P.
XX
XX 22-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-AUG-2000; 2000US-0228924P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.

Mon May 3 09:19:46 2004

us-10-603-260-1.rng

Page 17

Search completed: April 30, 2004, 11:04:10
Job time : 1064.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:27 ; Search time 7318.21 Seconds
(without alignments)
10519.601 Million cell updates/sec

Title: US-10-603-260-1
Perfect score: 2578
Sequence: 1 agctgcactatcacgccca.....tgataataagttctcttag 2578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estropl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esthum:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_pro:*
26: em_gss_pig:*
27: em_gss_vit:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 227 | 8.8 | 495 | 29 | CC818374 100004B07 |
| 2 | 225.4 | 8.7 | 496 | 29 | CC818523 100004L13 |
| 3 | 225.4 | 8.7 | 503 | 29 | CC818524 100006N08 |
| 4 | 224.8 | 8.7 | 308 | 14 | CF099977 rv37e09.y |

| | | | | | |
|----|-------|-----|-----|----|--------------------|
| 5 | 224.8 | 8.7 | 354 | 29 | CC818574 100004E20 |
| 6 | 224.8 | 8.7 | 374 | 29 | CC820036 100006M24 |
| 7 | 224.8 | 8.7 | 414 | 29 | CC819240 100005D19 |
| 8 | 224.8 | 8.7 | 464 | 29 | CC816935 100002C08 |
| 9 | 224.8 | 8.7 | 471 | 29 | CC817727 100003007 |
| 10 | 224.8 | 8.7 | 491 | 29 | CC819923 100006J13 |
| 11 | 224.8 | 8.7 | 507 | 9 | AJ568083 |
| 12 | 224.8 | 8.7 | 515 | 29 | CC817752 100003C16 |
| 13 | 224.8 | 8.7 | 518 | 29 | CC817128 100002D21 |
| 14 | 224.8 | 8.7 | 519 | 29 | CC817112 100002A23 |
| 15 | 224.8 | 8.7 | 519 | 29 | CC817796 100003K14 |
| 16 | 224.8 | 8.7 | 521 | 29 | CC819067 100005C09 |
| 17 | 224.8 | 8.7 | 527 | 29 | CC819808 100006F09 |
| 18 | 224.8 | 8.7 | 533 | 29 | CC819841 100006L07 |
| 19 | 224.8 | 8.7 | 539 | 29 | CC818925 100004H03 |
| 20 | 224.8 | 8.7 | 541 | 29 | CC817082 100002L15 |
| 21 | 224.8 | 8.7 | 541 | 29 | CC819843 100006L09 |
| 22 | 224.8 | 8.7 | 542 | 29 | CC816892 100002L01 |
| 23 | 224.8 | 8.7 | 542 | 29 | CC819005 100005G01 |
| 24 | 224.8 | 8.7 | 548 | 29 | CC817171 100002K22 |
| 25 | 224.8 | 8.7 | 551 | 29 | CC816905 100002N02 |
| 26 | 224.8 | 8.7 | 554 | 29 | CC819058 100005A09 |
| 27 | 224.8 | 8.7 | 559 | 29 | CC817875 100003J21 |
| 28 | 224.8 | 8.7 | 563 | 29 | CC819270 100005G21 |
| 29 | 224.8 | 8.7 | 563 | 29 | CC817022 100002B14 |
| 30 | 224.8 | 8.7 | 566 | 29 | CC820024 100006K23 |
| 31 | 224.8 | 8.7 | 567 | 29 | CC817070 100002J15 |
| 32 | 224.8 | 8.7 | 568 | 29 | CC818640 100004P22 |
| 33 | 224.8 | 8.7 | 571 | 29 | CC818423 100004J12 |
| 34 | 224.8 | 8.7 | 580 | 29 | CC819098 100005I07 |
| 35 | 224.8 | 8.7 | 581 | 29 | TA228H02Q |
| 36 | 224.8 | 8.7 | 583 | 29 | CC817633 100003M06 |
| 37 | 224.8 | 8.7 | 583 | 29 | CC818436 100004M08 |
| 38 | 224.8 | 8.7 | 584 | 29 | CC819862 100006O11 |
| 39 | 224.8 | 8.7 | 584 | 29 | CC932926 ZMMBC054 |
| 40 | 224.8 | 8.7 | 586 | 29 | CC816883 100002J03 |
| 41 | 224.8 | 8.7 | 588 | 29 | CC817788 100003I18 |
| 42 | 224.8 | 8.7 | 588 | 29 | CC818340 100004K04 |
| 43 | 224.8 | 8.7 | 589 | 29 | CC817595 100003G03 |
| 44 | 224.8 | 8.7 | 589 | 29 | CC817674 100003E11 |
| 45 | 224.8 | 8.7 | 590 | 29 | CC819754 100006L06 |

ALIGNMENTS

RESULT 1
LOCUS CC818374 495 bp DNA linear GSS 17-JUL-2003
DEFINITION 100004B07R Oxytricha plasmid U06C10 library Sterkiella
histicomscorum genomic clone U06C10004B07 R, genomic survey
sequence.

ACCESSION CC818374 GI:32897661
VERSION CC818374
KEYWORDS GSS.

SOURCE Sterkiella histicomscorum (Oxytricha trifallax)
ORGANISM Sterkiella histicomscorum
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
Stichotrichidae; Oxytrichidae; Sterkiella.

REFERENCE 1 (bases 1 to 495)
Dunn, D., Doak, R., Herrick, G. and Weiss, R.

AUTHORS Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
TITLE Unpublished (2003)
JOURNAL COMMENT

CONTACT: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel.: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Plate: 0004 row: B column: 07

Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 495.
 Location/Qualifiers
 1. 495

FEATURES
 source
 /organism="Sterkiella histriomuscorum"
 /mol_type="genomic DNA"
 /db_xref="taxon:94289"
 /clone="UUGC10004B07"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G14732114|9B|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
 Query Match 8.8%; Score 227; DB 29; Length 495;
 Best Local Similarity 94.0%; Pred. No. 1.2e-58;
 Matches 236; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2328 AAATGGCAAGTGTCTGCGGATACGCTGATGCGCATATGTTAAGCCAGCCCGACACC 2387
 DB 42 ATATGGGACATCTCAGACATCTGCTGATGCCCATATGTTAAGCCAGCCCGACACC 101
 QY 2388 CGGCAACACCCGCTGAGCGCCCTGACGGGCTGTGCTCCCGGATCCGCTTACAGAC 2447
 DB 102 CGCAACACCCGCTGAGCGCCCTGACGGGCTGTGCTCCCGGATCCGCTTACAGAC 161
 QY 2448 AAGCTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATACCGGAAC 2507
 DB 162 AAGCTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATACCGGAAC 221
 QY 2508 GCGGAGACGAAAGGGCTGTGATAGCCTATTTTATAGTTATGATGATATATA 2567
 DB 222 GCGGAGACGAAAGGGCTGTGATAGCCTATTTTATAGTTATGATGATATATA 281
 QY 2568 TGGTTCTTAG 2578
 DB 282 TGGTTCTTAG 292

RESULT 2 496 bp DNA linear GSS 17-JUL-2003
 CC818523
 LOCUS 100004113 R, genomic survey
 DEFINITION histriomuscorum genomic clone UUGC100004113 R, genomic survey
 sequence.
 ACCSSION CC818523.1 GI:32897943
 VERSION CC818523.1
 KEYWORDS GSS.
 SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)
 ORGANISM Sterkiella histriomuscorum
 Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
 Stichotrichida; Oxytrichidae; Sterkiella.
 1 (bases 1 to 496)
 REFERENCE 1
 AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.
 TITLE Paired end reads from plasmid inserts of Oxytricha trifallax
 JOURNAL macronuclear chromosomes
 COMMENT Unpublished (2003)
 CONTACT: Robert B. Weiss
 UNIVERSITY of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA

Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Plate: 0004 row: L column: 13
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 496.
 Location/Qualifiers
 1. 496

FEATURES
 source
 /organism="Sterkiella histriomuscorum"
 /mol_type="genomic DNA"
 /db_xref="taxon:94289"
 /clone="UUGC10004B07"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G14732114|9B|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 225.4; DB 29; Length 496;
 Best Local Similarity 93.6%; Pred. No. 3.7e-58;
 Matches 235; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2328 AAATGGCAAGTGTCTGCGGATACGCTGATGCGCATATGTTAAGCCAGCCCGACACC 2387
 DB 41 ATATGGGACATCTCAGACATCTGCTGATGCCCATATGTTAAGCCAGCCCGACACC 100
 QY 2388 CGGCAACACCCGCTGAGCGCCCTGACGGGCTGTGCTCCCGGATCCGCTTACAGAC 2447
 DB 101 CGCAACACCCGCTGAGCGCCCTGACGGGCTGTGCTCCCGGATCCGCTTACAGAC 160
 QY 2448 AAGCTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATACCGGAAC 2507
 DB 161 AAGCTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATACCGGAAC 220
 QY 2508 GCGGAGACGAAAGGGCTGTGATAGCCTATTTTATAGTTATGATGATATATA 2567
 DB 221 GCGGAGACGAAAGGGCTGTGATAGCCTATTTTATAGTTATGATGATATATA 280
 QY 2568 TGGTTCTTAG 2578
 DB 281 TGGTTCTTAG 291

RESULT 3 503 bp DNA linear GSS 17-JUL-2003
 CC819854
 LOCUS 100006088 R, genomic survey
 DEFINITION histriomuscorum genomic clone UUGC100006088 R, genomic survey
 sequence.
 ACCSSION CC819854.1 GI:32900533
 VERSION CC819854.1
 KEYWORDS GSS.
 SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)
 ORGANISM Sterkiella histriomuscorum
 Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
 Stichotrichida; Oxytrichidae; Sterkiella.
 1 (bases 1 to 503)
 REFERENCE 1
 AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.
 TITLE Paired end reads from plasmid inserts of Oxytricha trifallax
 JOURNAL macronuclear chromosomes
 COMMENT Unpublished (2003)
 CONTACT: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 309, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Place: 0006 row: N column: 08
Seq primer: CACACAGGAACGCTTGTGACC
class: plasmid ends
high quality sequence stop: 503.

ORIGIN

Query Match 8.7%; Score 225.4; DB 29; Length 503;

| | | | |
|----|------|--|------|
| Qy | 2328 | AATTTGGCAAGTGTCTGCCGCGAATACGCTGAATGCCGATAGTTAAAGCCAGCCCCGACACC | 2387 |
| Db | 60 | ATTATGTGCATCCTCATGTAACAATCGCTCTGTATGCCGATAGTTAAAGCCAGCCCCGACACC | 119 |
| Qy | 2388 | CGCCAACACCCGCTGACGCGCCCTGACAGGGAGTTGTCTGCTCCCGGCAATCCGCTTAACAGAC | 2447 |
| Db | 120 | CGCACAAACCCGCTACAGCGCCCTGAACGGAGCTTGTCGTCTCCGCGAATCCGCTTAACAGAC | 179 |
| Qy | 2448 | AAGCTGTACCGTCTTCGCGAGCTGCATGTGTCAAGAGTTTQAACGTCATCACCGAAAC | 2507 |
| Db | 180 | AACCTGTGACCTCTCCCGGAGCTGCATGTGTCAAGAGTTTCAACGTCATCACCGAAAC | 239 |
| Qy | 2508 | GCGCGAGAAGAAAGGCGCTCGTAGATACGCTATTATTTATAGGTTAATGTCATGATAATA | 2567 |
| Db | 240 | GCGCGAGAAGAAAGGCGCTCGTAGATACGCTATTATTTATAGGTTAATGTCATGATAATA | 299 |
| Qy | 2568 | TGGTTTTCTTAG | 2578 |
| Db | 300 | TGGTTTTCTTAG | 310 |

RESULT 4

| | | | | | |
|----------|------------|------------|---|-----------------------------|-------------------------------|
| CF099977 | LOCUS | 308 bp | mRNA | linear | EST 22-UTL-2003 |
| CF099977 | DEFINITION | rv37609_y1 | <i>Heterodera schachtii</i> J2 PAMP1 v1 | <i>Heterodera schachtii</i> | |
| | | CDNA 5' | similar to TR:Q51617 | Q51617 | 7.6 kD ORF. ;, mRNA sequence. |

REFERENCE
AUTHORS
1 (bases 1 to 308)
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J.,
Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,

| TITLE | JOURNAL | COMMENT |
|---|---------|---------|
| Unpublished (1999) | | |
| The Washington Univ. Nematode ESP Project, 1999 | | |
| Contact: McCarter JP | | |
| Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsaferashvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurr, R., Kohn, S., Shint, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R. | | |

FEATURES

Source

ORIGIN

| | | | | |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match | 8.7% | Score 224.8; | DB 14; | Length 308; |
| Best Local Similarity | 99.1%; | Pred. No. 4.2e-58; | | |
| Matches 226; | Conservative 0; | Mismatches 2; | Indels 0; | Gaps 0; |

| QY | 2351 | TACGGTANTAGCCGAGATGTTAAAGCAGAGCCCCGACACCCGCGCAACACCGGTACAGGGCCC | 2411 |
|----|------|---|------|
| Db | 78 | TGCTTGATAGCCGAGATGTTAAGCAGAGCCCCGACACCCGCGCAACACCGGTACAGGGCCC | 137 |
| QY | 2411 | TGACGGGGCTTGTCTGCTCCGGGATCCGGCTTACAGACAAGCTGTGACCGTCTCCGGGAGAC | 2470 |
| Db | 138 | TGACGGGGCTTGTCTGCTCCGGGATCCGGCTTACAGACAAGCTGTGACCGTCTCCGGGAGAC | 197 |
| QY | 2471 | TGCATGTGTCAAGAGTTTTCACCGTCAATCACCGAAACGGCGGACAGAAAGGGCCTCGTG | 2533 |
| Db | 198 | TGCATGTGTCAAGAGTTTTCACCGTCAATCACCGAAACGGCGGACAGAAAGGGCCTCGTG | 257 |
| QY | 2531 | TACGCCCTATTTTATAGGTTAATGCATCATATATATATAGTTCCTTAG | 2578 |
| Db | 258 | TACGCCCTATTTTATAGGTTAATGCATCATATATATATAGTTCCTTAG | 305 |

RESULT 5

| CC818574 | LOCUS | CC818574 | 364 bp | DNA | linear | GSS 17-JUL-2003 |
|------------|-------|-----------------|---------------|-----------------|----------------|-----------------|
| DEFINITION | | 100004220R | Oxytricha | plasmid | U0C10 library | Sterkiella |
| | | histrionuscorum | genomic clone | U0C100004220 R, | genomic survey | |

SOURCE
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE
AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.
TITLE Paired end reads from plasmid inserts of Oxytricha trifallax
1 (bases 1 to 364)

JOURNAL
COMMENT Unpublished (2003)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0004 row: E column: 20
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 364.
Location/Qualifiers

FEATURES

source
1. .364
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="U000000420"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G114732114|G1|AR129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 364;
Best Local Similarity 99.1%; Pred. No. 4,7e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCGGATAGTTAAGCCAGCCCGGACACCCGCTGACGGGCC 2410
DB 87 TGCTGTATGCGGATAGTTAAGCCAGCCCGGACACCCGCTGACGGGCC 146
QY 2411 TGAGGGGCTGTGCTGCCGCGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 2470
DB 147 TGACGGGCTGTGCTGCCGCGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 206
QY 2471 TGCATGTGTGAGAGTTTTCACCGCTATCAACCGAAGCGCGAGAGAGGCTCTG 2530
DB 207 TGCATGTGTGAGAGTTTTCACCGCTATCAACCGAAGCGCGAGAGAGGCTCTG 266
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTAG 2578
DB 267 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTAG 314

RESULT 6
CC820036 374 bp DNA linear GSS 17-JUL-2003
LOCUS CC820036
DEFINITION 100005M24R Oxytricha plasmid U0000 library Sterkiella
histriomuscorum genomic clone U0000006M24 R, genomic survey
sequence.
ACCESSION CC820036
VERSION CC820036.1 GI:32900898
KEYWORDS GSS.

SOURCE
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE
AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.
TITLE Paired end reads from plasmid inserts of Oxytricha trifallax
1 (bases 1 to 374)

JOURNAL
COMMENT Unpublished (2003)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0006 row: M column: 24
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 374.
Location/Qualifiers

FEATURES

source
1. .374
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="U0000006M24"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G114732114|G1|AR129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 374;
Best Local Similarity 99.1%; Pred. No. 4.8e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCGGATAGTTAAGCCAGCCCGGACACCCGCTGACGGGCC 2410
DB 92 TGCTGTATGCGGATAGTTAAGCCAGCCCGGACACCCGCTGACGGGCC 151
QY 2411 TGAGGGGCTGTGCTGCCGCGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 2470
DB 152 TGACGGGCTGTGCTGCCGCGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 211
QY 2471 TGCATGTGTGAGAGTTTTCACCGCTATCAACCGAAGCGCGAGAGAGGCTCTG 2530
DB 212 TGCATGTGTGAGAGTTTTCACCGCTATCAACCGAAGCGCGAGAGAGGCTCTG 271
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTAG 2578
DB 272 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTAG 319

RESULT 7
CC819240 414 bp DNA linear GSS 17-JUL-2003
LOCUS CC819240
DEFINITION 100005D19R Oxytricha plasmid U0000 library Sterkiella
histriomuscorum genomic clone U0000005D19 R, genomic survey
sequence.
ACCESSION CC819240
VERSION CC819240.1 GI:32899308
KEYWORDS GSS.

SOURCE
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE
AUTHORS Sterkiella histriomuscorum
Stichotrichida; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
1 (bases 1 to 414)

TITLE
Dunn, D., Doak, T., Herrick, G. and Weiss, R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)

JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., StC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0005 row: D column: 19
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 414.

FEATURES
source
1..414
Location/Qualifiers
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUGC100005D19"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUGC10 library"
/note="Vector: pMD22ny; Purified macronuclear chromosomal
DNA from Oxytricha trifallax was blunt end-repaired with
T4 DNA polymerase and T4 polynucleotide kinase. Adaptor
oligonucleotides were ligated to the blunt ends in high
molar excess. Vector DNA was prepared from a derivative of
pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible
derivative of plasmid R1. The vector was ligated with
adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. Coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 8.7%; Score 224.8; DB 29; Length 414;
Best Local Similarity 99.1%; Pred. No. 5.1e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TAGCGGATGCGGATGTTAGCCAGCCCGCAACCCGCTGAGCGGCC 2410
DB 87 TGCTCTGATGCGGATGTTAGCCAGCCCGCAACCCGCTGAGCGGCC 146
QY 2411 TGACGGGCTTGTCTGCTCCGGGATCCGCTTACAGACAACTGTGACCGTCTCCGGGAGC 2470
DB 147 TGACGGGCTTGTCTGCTCCGGGATCCGCTTACAGACAACTGTGACCGTCTCCGGGAGC 206
QY 2471 TGATGTGTGAGAGGTTTACCGGTATACCGAAACGGCGAGACGAAAGGCGCTCTG 2530
DB 207 TGATGTGTGAGAGGTTTACCGGTATACCGAAACGGCGAGACGAAAGGCGCTCTG 266
QY 2531 ATACGCTATTTTATAGGTTATGTCATGATATATATAGTTCTTAG 2578
DB 267 ATACGCTATTTTATAGGTTATGTCATGATATATATAGTTCTTAG 314

RESULT 8
CC816935 464 bp DNA linear GSS 17-JUL-2003
LOCUS 1000020C08R Oxytricha plasmid UUGC10 library Sterkiella
DEFINITION histriomuscorum genomic clone UUGC1000020C08 R, genomic survey
sequence.
ACCESSION CC816935
VERSION CC816935.1 GI:32896222
KEYWORDS GSS.

SOURCE
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE
AUTHORS Sterkiella histriomuscorum
Stichotrichida; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
1 (bases 1 to 464)

TITLE
Dunn, D., Doak, T., Herrick, G. and Weiss, R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)

JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., StC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0002 row: C column: 08
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 464.

FEATURES
source
1..464
Location/Qualifiers
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUGC100002C08"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUGC10 library"
/note="Vector: pMD22ny; Purified macronuclear chromosomal
DNA from Oxytricha trifallax was blunt end-repaired with
T4 DNA polymerase and T4 polynucleotide kinase. Adaptor
oligonucleotides were ligated to the blunt ends in high
molar excess. Vector DNA was prepared from a derivative of
pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible
derivative of plasmid R1. The vector was ligated with
adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. Coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 8.7%; Score 224.8; DB 29; Length 464;
Best Local Similarity 99.1%; Pred. No. 5.4e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TAGCGGATGCGGATGTTAGCCAGCCCGCAACCCGCTGAGCGGCC 2410
DB 180 TGCTCTGATGCGGATGTTAGCCAGCCCGCAACCCGCTGAGCGGCC 239
QY 2411 TGACGGGCTTGTCTGCTCCGGGATCCGCTTACAGACAACTGTGACCGTCTCCGGGAGC 2470
DB 240 TGACGGGCTTGTCTGCTCCGGGATCCGCTTACAGACAACTGTGACCGTCTCCGGGAGC 299
QY 2471 TGATGTGTGAGAGGTTTACCGGTATACCGAAACGGCGAGACGAAAGGCGCTCTG 2530
DB 300 TGATGTGTGAGAGGTTTACCGGTATACCGAAACGGCGAGACGAAAGGCGCTCTG 359
QY 2531 ATACGCTATTTTATAGGTTATGTCATGATATATATAGTTCTTAG 2578
DB 360 ATACGCTATTTTATAGGTTATGTCATGATATATATAGTTCTTAG 407

RESULT 9
CC817727 471 bp DNA linear GSS 17-JUL-2003
LOCUS 100003007R Oxytricha plasmid UUGC10 library Sterkiella
DEFINITION histriomuscorum genomic clone UUGC100003007 R, genomic survey
sequence.
ACCESSION CC817727
VERSION CC817727.1 GI:32897014
KEYWORDS GSS.

SOURCE
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE
AUTHORS Sterkiella histriomuscorum
TITLE Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
1 (bases 1 to 471)
Dunn, D., Doak, T., Herrick, G. and Weiss, R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)
JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0003 row: 0 column: 07
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 471.
Location/Qualifiers

FEATURES
source
1..471
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUCG100003007"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUCG10 library"
/note="Vector: pMD42nv; Purified macronuclear chromosomal
DNA from Oxytricha trifallax was blunt end-repaired with
T4 DNA polymerase and T4 polynucleotide kinase. Adaptor
oligonucleotides were ligated to the blunt ends in high
molar excess. Vector DNA was prepared from a derivative of
pMD42 (g114732114|gb|AF129072.1), a copy-number inducible
derivative of plasmid R1. The vector was ligated with
adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. Coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 8.7%; Score 224.8; DB 29; Length 471;
Best Local Similarity 99.1%; Pred. No. 5.5e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCATAGTTAAGCCAGCCCGCAACCCCGTGAAGCGGCC 2410
DB 91 TGCTGTGATGCGCATAGTTAAGCCAGCCCGCAACCCCGTGAAGCGGCC 150
QY 2411 TGACGGGGCTTGTCTGCTCCCGCATCGCTTACAGACAGCTGACCGTCCGGAGC 2470
DB 151 TGACGGGGCTTGTCTGCTCCCGCATCGCTTACAGACAGCTGACCGTCCGGAGC 210
QY 2471 TGCATGTGTCAAGAGTTTTCACCGTCATCAGCGAAAGCGCGAGACGAAAGGCGCTGTG 2530
DB 211 TGCATGTGTCAAGAGTTTTCACCGTCATCAGCGAAAGCGCGAGACGAAAGGCGCTGTG 270
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTTAG 2578
DB 271 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTTAG 318

RESULT 10
CC819923 491 bp DNA linear GSS 17-JUN-2003
LOCUS CC819923
DEFINITION 1000060613r Oxytricha plasmid UUCG10 library Sterkiella
histriomuscorum genomic clone UUCG1000060613 R, genomic survey
sequence.
ACCESSION CC819923
VERSION CC819923.1 GI:33290671
KEYWORDS GSS.

SOURCE
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE
AUTHORS Sterkiella histriomuscorum
TITLE Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
1 (bases 1 to 491)
Dunn, D., Doak, T., Herrick, G. and Weiss, R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)
JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0006 row: 0 column: 13
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 491.
Location/Qualifiers

FEATURES
source
1..491
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUCG1000060613"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUCG10 library"
/note="Vector: pMD42nv; Purified macronuclear chromosomal
DNA from Oxytricha trifallax was blunt end-repaired with
T4 DNA polymerase and T4 polynucleotide kinase. Adaptor
oligonucleotides were ligated to the blunt ends in high
molar excess. Vector DNA was prepared from a derivative of
pMD42 (g114732114|gb|AF129072.1), a copy-number inducible
derivative of plasmid R1. The vector was ligated with
adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. Coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 8.7%; Score 224.8; DB 29; Length 491;
Best Local Similarity 99.1%; Pred. No. 5.6e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCATAGTTAAGCCAGCCCGCAACCCCGTGAAGCGGCC 2410
DB 85 TGCTGTGATGCGCATAGTTAAGCCAGCCCGCAACCCCGTGAAGCGGCC 144
QY 2411 TGACGGGGCTTGTCTGCTCCCGCATCGCTTACAGACAGCTGACCGTCCGGAGC 2470
DB 145 TGACGGGGCTTGTCTGCTCCCGCATCGCTTACAGACAGCTGACCGTCCGGAGC 204
QY 2471 TGCATGTGTCAAGAGTTTTCACCGTCATCAGCGAAAGCGCGAGACGAAAGGCGCTGTG 2530
DB 205 TGCATGTGTCAAGAGTTTTCACCGTCATCAGCGAAAGCGCGAGACGAAAGGCGCTGTG 264
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTTAG 2578
DB 265 ATAGCGCTATTTTATAGTTAATGTCATGATTAATAGTTTCTTAG 312

RESULT 11
AJ568083 507 bp mRNA linear EST 28-JUN-2003
LOCUS AJ568083
DEFINITION CDNA clone 2835, mRNA sequence.
ACCESSION AJ568083
VERSION AJ568083.1 GI:33293972
KEYWORDS EST
Antirrhinum majus (snapdragon)

ORGANISM
Antirrhinum majus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Lamiaceae; Antirrhinaceae; Antirrhineae;
Antirrhinum.
1 (bases 1 to 507)
Stueber, K. and Schwarz-Sommer, Z.
Antirrhinum majus EST collection (Stueber, K. and Schwarz-Sommer, Z.)
JOURNAL
Unpublished (2003)
COMMENT
Contact: Stueber K
Molecular Plant Genetics
MPI fuer Zuechtungsforchung
Carl-von-Linne Weg 10, D-50829, Germany.
Location/Qualifiers
1. 507
/organism="Antirrhinum majus"
/mol_type="mRNA"
/db_xref="taxon:4151"
/clone="2835"
/clone_lib="Antirrhinum majus library (Stueber K)"

ORIGIN
Query Match 8.7%; Score 224.8; DB 9; Length 507;
Best Local Similarity 99.1%; Pred. No. 5.7e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
2351 TAGCGTGAATGCGCATGTTAAGCCAGCCCGGACACCCGCAACACCCGCTGACGCGCC 2410
273 TGCTGTATGCGCATGTTAAGCCAGCCCGGACACCCGCAACACCCGCTGACGCGCC 332
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 2470
DB 333 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 392
QY 2471 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTGTG 2530
DB 393 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTGTG 452
QY 2531 ATACGCCATATTTTATAGTTAATGTCATGATATATATATGTTCTTAG 2578
DB 453 ATACGCCATATTTTATAGTTAATGTCATGATATATATATGTTCTTAG 500

RESULT 12
CC817752 515 bp DNA linear GSS 17-JUL-2003
LOCUS
DEFINITION
CC817752 100003C16R Oxytricha plasmid UUGC10 library Sterkiella
histrionuscorum genomic clone UUGC10003C16 R, genomic survey
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.
Sterkiella histrionuscorum (Oxytricha trifallax)
Sterkiella histrionuscorum
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
Stichotrichida; Oxytrichidae; Sterkiella.
1 (bases 1 to 515)
Dunn, D., Doak, T., Herrick, G. and Weiss, R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0003 row: C column: 16
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 515.

FEATURES
source
Location/Qualifiers
1. 515
/organism="Sterkiella histrionuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUGC10003C16"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUGC10 library"
/note="Vector: PMD42nv; Purified macronuclear chromosomal
DNA from Oxytricha trifallax was blunt end-repaired with
T4 DNA polymerase and T4 polynucleotide kinase. Adaptor
oligonucleotides were ligated to the blunt ends in high
molar excess. Vector DNA was prepared from a derivative of
pMD42 (g14732114|gb|AF129072.1), a copy-number inducible
derivative of plasmid R1. The vector was ligated with
adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. Coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 8.7%; Score 224.8; DB 29; Length 515;
Best Local Similarity 99.1%; Pred. No. 5.8e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
2351 TAGCGTGAATGCGCATGTTAAGCCAGCCCGGACACCCGCAACACCCGCTGACGCGCC 2410
85 TGCTGTATGCGCATGTTAAGCCAGCCCGGACACCCGCAACACCCGCTGACGCGCC 144
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 2470
DB 145 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 204
QY 2471 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTGTG 2530
DB 205 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTGTG 264
QY 2531 ATACGCCATATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578
DB 265 ATACGCCATATTTTATAGTTAATGTCATGATATATATGTTCTTAG 312

RESULT 13
CC817128 518 bp DNA linear GSS 17-JUL-2003
LOCUS
DEFINITION
CC817128 100002D21R Oxytricha plasmid UUGC10 library Sterkiella
histrionuscorum genomic clone UUGC10002D21 R, genomic survey
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.
Sterkiella histrionuscorum (Oxytricha trifallax)
Sterkiella histrionuscorum
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
Stichotrichida; Oxytrichidae; Sterkiella.
1 (bases 1 to 518)
Dunn, D., Doak, T., Herrick, G. and Weiss, R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Plate: 0002 row: D column: 21
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 518.

FEATURES
source

Location/Qualifiers
1. 518
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUGC10002A23"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUGC10 library"
/note="Vector: PMD42nv. Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (g14732114[gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 518;
Best Local Similarity 99.1%; Pred. No. 5.8e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2351 TAGCGTATGCGCATAGTTAAGCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 2410
83 TGCTGTATGCGCATAGTTAAGCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 142
2411 TGACGGGCTTGTCTGCTCCGGCATCGCTTAAGAGACAGCTGTACCGGAGC 2470
143 TGACGGGCTTGTCTGCTCCGGCATCGCTTAAGAGACAGCTGTACCGGAGC 202
2471 TGATGTGTGACAGAGTTTCAACCGTCATCACCGGAGAGAGGAGGAGGAGG 2530
203 TGATGTGTGACAGAGTTTCAACCGTCATCACCGGAGAGAGGAGGAGGAGG 262
2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATATGTTCTTAG 2578
263 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 310

RESULT 14
CC817112 519 bp DNA linear GSS 17-JUL-2003
LOCUS 100002A23R Oxytricha plasmid UUGC10 library Sterkiella
DEFINITION histriomuscorum genomic clone UUGC100002A23 R, genomic survey
sequence.
ACCESSION CC817112
VERSION CC817112.1 GI:32896399
KEYWORDS
SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)
ORGANISM Sterkiella histriomuscorum
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
Stichotrichida; Oxytrichidae; Sterkiella.
1 (bases 1 to 519)
Dunn,D., Doak,T., Herrick,G. and Weiss,R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)
CONTACT: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Plate: 0002 row: A column: 23
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 519.

FEATURES
source

Location/Qualifiers
1. 519
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUGC10002A23"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Oxytricha plasmid UUGC10 library"
/note="Vector: PMD42nv. Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (g14732114[gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 519;
Best Local Similarity 99.1%; Pred. No. 5.8e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2351 TAGCGTATGCGCATAGTTAAGCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 2410
225 TGCTGTATGCGCATAGTTAAGCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 294
2411 TGACGGGCTTGTCTGCTCCGGCATCGCTTAAGAGACAGCTGTACCGGAGC 2470
225 TGACGGGCTTGTCTGCTCCGGCATCGCTTAAGAGACAGCTGTACCGGAGC 354
2471 TGATGTGTGACAGAGTTTCAACCGTCATCACCGGAGAGAGGAGGAGGAGG 2530
355 TGATGTGTGACAGAGTTTCAACCGTCATCACCGGAGAGAGGAGGAGGAGG 414
2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATATGTTCTTAG 2578
415 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 462

RESULT 15
CC817796 519 bp DNA linear GSS 17-JUL-2003
LOCUS 100003K14R Oxytricha plasmid UUGC10 library Sterkiella
DEFINITION histriomuscorum genomic clone UUGC10003K14 R, genomic survey
sequence.
ACCESSION CC817796
VERSION CC817796.1 GI:32897083
KEYWORDS
SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)
ORGANISM Sterkiella histriomuscorum
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
Stichotrichida; Oxytrichidae; Sterkiella.
1 (bases 1 to 519)
Dunn,D., Doak,T., Herrick,G. and Weiss,R.
Paired end reads from plasmid inserts of Oxytricha trifallax
macronuclear chromosomes
Unpublished (2003)
CONTACT: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Plate: 0003 row: K column: 14
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 519.

FEATURES
source

Location/Qualifiers
1. 519
/organism="Sterkiella histriomuscorum"
/mol_type="genomic DNA"
/db_xref="taxon:94289"
/clone="UUGC100003X14"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Oxytricha plasmid UUGC10 library"
/note="Vector: pMD42ny; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 519;
Best Local Similarity 99.1%; Pred. No. 5.8e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TACGCTGATGCCGATAGTTAAGCCAGCCCGCAACCCCGCAACCCGCTGACGGCCC 2410
DB 88 TGCTTGATGCCGATAGTTAAGCCAGCCCGCAACCCCGCAACCCGCTGACGGCCC 147
QY 2411 TGACGGGCTTGTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGCTCCGGAGC 2470
DB 148 TGACGGGCTTGTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGCTCCGGAGC 207
QY 2471 TGATGTGACAGAGTTTCAACCGTATGACGAAACGGCGAGAGGAAAGGCGCTCGTG 2530
DB 208 TGATGTGACAGAGTTTCAACCGTATGACGAAACGGCGAGAGGAAAGGCGCTCGTG 267
QY 2531 ATACGCTATTTTATAGGTATATGATGATATATATGTTCTTAG 2578
DB 268 ATACGCTATTTTATAGGTATATGATGATATATATGTTCTTAG 315

Search completed: April 30, 2004, 18:34:55
Job time : 7322.21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:32 ; Search time 302.986 Seconds
(without alignments)
4721.887 Million cell updates/sec

Title: US-10-603-260-1

Perfect score: 2578

Sequence: 1 agcttcacattatcagcca.....tgataataatggtttcttag 2578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 224.8 | 8.7 | 1073 | 4 | US-09-481-049-6 |
| 2 | 224.8 | 8.7 | 2580 | 4 | US-09-194-285-23 |
| 3 | 224.8 | 8.7 | 2939 | 1 | US-08-119-512-2 |
| 4 | 224.8 | 8.7 | 2939 | 1 | US-08-488-015B-2 |
| 5 | 224.8 | 8.7 | 2939 | 3 | US-08-814-412-17 |
| 6 | 224.8 | 8.7 | 3104 | 1 | US-07-415-307A-1 |
| 7 | 224.8 | 8.7 | 3104 | 1 | US-08-371-320-1 |
| 8 | 224.8 | 8.7 | 3190 | 4 | US-09-027-169-6 |
| 9 | 224.8 | 8.7 | 3343 | 6 | 5453363-2 |
| 10 | 224.8 | 8.7 | 3796 | 4 | US-09-470-661A-32 |
| 11 | 224.8 | 8.7 | 3799 | 4 | US-09-173-053-4 |
| 12 | 224.8 | 8.7 | 3819 | 3 | US-09-042-353-393 |
| 13 | 224.8 | 8.7 | 3819 | 4 | US-08-758-417A-243 |
| 14 | 224.8 | 8.7 | 3875 | 3 | US-09-039-982A-21 |
| 15 | 224.8 | 8.7 | 3875 | 3 | US-09-039-641-21 |
| 16 | 224.8 | 8.7 | 3875 | 3 | US-09-039-762A-21 |
| 17 | 224.8 | 8.7 | 3875 | 4 | US-09-042-492D-21 |
| 18 | 224.8 | 8.7 | 3875 | 4 | US-08-913-612A-21 |
| 19 | 224.8 | 8.7 | 3878 | 3 | US-09-039-982A-27 |
| 20 | 224.8 | 8.7 | 3878 | 3 | US-09-039-641-27 |
| 21 | 224.8 | 8.7 | 3878 | 4 | US-09-042-492D-27 |
| 22 | 224.8 | 8.7 | 3878 | 4 | US-08-913-612A-27 |
| 23 | 224.8 | 8.7 | 3878 | 4 | US-09-042-492D-27 |
| 24 | 224.8 | 8.7 | 3881 | 3 | US-09-042-353-369 |
| 25 | 224.8 | 8.7 | 3881 | 4 | US-08-758-417A-217 |
| 26 | 224.8 | 8.7 | 3883 | 4 | US-09-039-982A-30 |
| 27 | 224.8 | 8.7 | 3883 | 3 | US-09-039-641-30 |

| | | | | | | |
|------|-------|-----|------|---|-------------------|-------------------|
| 28 | 224.8 | 8.7 | 3883 | 3 | US-09-039-762A-30 | Sequence 30, Appl |
| 29 | 224.8 | 8.7 | 3883 | 4 | US-09-042-492D-30 | Sequence 30, Appl |
| 30 | 224.8 | 8.7 | 3883 | 4 | US-08-913-612A-30 | Sequence 30, Appl |
| 31 | 224.8 | 8.7 | 3908 | 3 | US-09-039-982A-24 | Sequence 24, Appl |
| 32 | 224.8 | 8.7 | 3908 | 3 | US-09-039-641-24 | Sequence 24, Appl |
| 33 | 224.8 | 8.7 | 3908 | 3 | US-09-039-762A-24 | Sequence 24, Appl |
| 34 | 224.8 | 8.7 | 3908 | 4 | US-09-042-492D-24 | Sequence 24, Appl |
| 35 | 224.8 | 8.7 | 3908 | 4 | US-08-913-612A-24 | Sequence 24, Appl |
| 36 | 224.8 | 8.7 | 4016 | 1 | US-08-040-753-1 | Sequence 1, Appl |
| 37 | 224.8 | 8.7 | 4016 | 4 | US-09-173-053-3 | Sequence 3, Appl |
| 38 | 224.8 | 8.7 | 4045 | 3 | US-08-464-700-54 | Sequence 54, Appl |
| 39 | 224.8 | 8.7 | 4054 | 3 | US-09-098-287A-9 | Sequence 9, Appl |
| 40 | 224.8 | 8.7 | 4054 | 4 | US-09-445-649-9 | Sequence 9, Appl |
| C 41 | 224.8 | 8.7 | 4118 | 4 | US-09-068-821-17 | Sequence 17, Appl |
| C 42 | 224.8 | 8.7 | 4118 | 4 | US-09-068-821-18 | Sequence 18, Appl |
| C 43 | 224.8 | 8.7 | 4161 | 4 | US-09-185-244-8 | Sequence 8, Appl |
| C 44 | 224.8 | 8.7 | 4161 | 4 | US-09-171-913-1 | Sequence 1, Appl |
| C 45 | 224.8 | 8.7 | 4273 | 3 | US-08-795-430-3 | Sequence 3, Appl |

ALIGNMENTS

RESULT 1
US-09-481-049-6/c
Sequence 6, Application US/09481049
Patent No. 6395485
GENERAL INFORMATION:
APPLICANT: DE BEUCKELER, Marc
TITLE OF INVENTION: METHODS FOR IDENTIFYING ELITE EVENT GAT-ZM1 IN
FILE REFERENCE: 514412-2025
CURRENT APPLICATION NUMBER: US/09/481,049
CURRENT FILING DATE: 2000-01-11
NUMBER OF SEQ. ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1073
TYPE: DNA
ORGANISM: Zea mays
US-09-481-049-6

Query Match 8.7%; Score 224.8; DB 4; Length 1073;
Best Local Similarity 99.1%; Pred. No. 7e-60; 2; Indels 0; Gaps 0;
Matches 226; Conservative 0; Mismatches 2

QY 2351 TAGCTGATGCCGATAGTTAAGCCAGCCCGGCAACCCGCTGACGGGCC 2410
DB 709 TGCTGTGATGCCGATAGTTAAGCCAGCCCGGCAACCCGCTGACGGGCC 650
QY 2411 TGACGGGCTGTGCTCCCGGATCCGCTTACAGCAAGCTGTACCGTCCGGGAGC 2470
DB 649 TGACGGGCTGTGCTCCCGGATCCGCTTACAGCAAGCTGTACCGTCCGGGAGC 590
QY 2471 TGATGTGTGAGAGTTTTCACCGTATCATCCGATACCGAAGCGCGAGAGAGGCGCTGTG 2530
DB 589 TGATGTGTGAGAGTTTTCACCGTATCATCCGATACCGAAGCGCGAGAGAGGCGCTGTG 530
QY 2531 ATAGCGCTATTTTATAGGTTAATGTCATGATTAATATGTTCTTAG 2578
DB 529 ATAGCGCTATTTTATAGGTTAATGTCATGATTAATATGTTCTTAG 482

RESULT 2
US-09-194-285-23
Sequence 23, Application US/09194285
Patent No. 6355479
GENERAL INFORMATION:
APPLICANT: Webd, Susan R.
APPLICANT: Windyast, Oia
APPLICANT: Karlsson, Lars
APPLICANT: Jackson, Michael R.
APPLICANT: Peterson, Per A.

TITLE OF INVENTION: MEC Class II Antigen Presenting Systems
TITLE OF INVENTION: and Methods for Activating CD4+ T Cells
FILE REFERENCE: TSRI 536.1
CURRENT APPLICATION NUMBER: US/09/194,285
CURRENT FILING DATE: 1999-04-12
PRIORITY APPLICATION NUMBER: PCT/US97/08697
PRIORITY FILING DATE: 1997-05-22
PRIORITY APPLICATION NUMBER: US 60/018,175
PRIORITY FILING DATE: 1996-05-23
NUMBER OF SEQ ID NOS: 56
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 2580
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: engineered recombinant
US-09-194-285-23

Query Match 8.7%; Score 224.8; DB 4; Length 2580;
Best Local Similarity 99.1%; Pred. No. 1.2e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCGCGATAGTTAAGCCAGCCCGGACACCCCGGACACCCGCTGAGCGGCC 2410
DB 2126 TGCTCTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGGACACCCGCTGAGCGGCC 2185
QY 2411 TGACGGGCTGTGCTGCTCCCGGATCCGCTTACAGACAGAGCTGTGACCGCTTCCGGAGC 2470
DB 2186 TGACGGGCTGTGCTGCTCCCGGATCCGCTTACAGACAGAGCTGTGACCGCTTCCGGAGC 2245
QY 2471 TGATGTGTGAGAGGTTTTCACCGCTATCAGACGAGGAGGAGGAGGAGGAGGAGGAGG 2530
DB 2246 TGATGTGTGAGAGGTTTTCACCGCTATCAGACGAGGAGGAGGAGGAGGAGGAGGAGG 2305
QY 2531 ATAGCGCTATTTTAAAGTTAATGTCATGATGATGATGATGATGATGATGATGATGATG 2578
DB 2306 ATAGCGCTATTTTAAAGTTAATGTCATGATGATGATGATGATGATGATGATGATGATG 2353

RESULT 3
US-08-119-512-2
Sequence 2, Application US/08119512
Patent No. 5498531
GENERAL INFORMATION:
APPLICANT: Jarrell, Kevin A.
TITLE OF INVENTION: INTRON MEDIATED RECOMBINANT TECHNIQUES
TITLE OF INVENTION: AND REAGENTS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,512
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HUT-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2939 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2448..2657
OTHER INFORMATION: /product= "b-globin exon 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2667..2814
OTHER INFORMATION: /product= "b-globin exon 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2815..2890
OTHER INFORMATION: /product= "intron sequence"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2390..2447
OTHER INFORMATION: /product= "intron sequence"
US-08-119-512-2

Query Match 8.7%; Score 224.8; DB 1; Length 2939;
Best Local Similarity 99.1%; Pred. No. 1.3e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCGCGATAGTTAAGCCAGCCCGGACACCCCGGACACCCGCTGAGCGGCC 2410
DB 105 TGCTCTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGGACACCCGCTGAGCGGCC 164
QY 2411 TGACGGGCTGTGCTGCTCCCGGATCCGCTTACAGACAGAGCTGTGACCGCTTCCGGAGC 2470
DB 165 TGACGGGCTGTGCTGCTCCCGGATCCGCTTACAGACAGAGCTGTGACCGCTTCCGGAGC 224
QY 2471 TGATGTGTGAGAGGTTTTCACCGCTATCAGACGAGGAGGAGGAGGAGGAGGAGGAGG 2530
DB 225 TGATGTGTGAGAGGTTTTCACCGCTATCAGACGAGGAGGAGGAGGAGGAGGAGGAGG 284
QY 2531 ATAGCGCTATTTTAAAGTTAATGTCATGATGATGATGATGATGATGATGATGATGATG 2578
DB 285 ATAGCGCTATTTTAAAGTTAATGTCATGATGATGATGATGATGATGATGATGATGATG 332

RESULT 4
US-08-488-015B-2
Sequence 2, Application US/08488015B
Patent No. 5780272
GENERAL INFORMATION:
APPLICANT: Jarrell, Kevin A.
TITLE OF INVENTION: INTRON-MEDIATED RECOMBINANT TECHNIQUES
TITLE OF INVENTION: AND REAGENTS
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Eliot
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,015B
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HUT-008.02

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2939 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2448..2657
OTHER INFORMATION: /product= "p-globin exon 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2667..2814
OTHER INFORMATION: /product= "p-globin exon 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2815..2890
OTHER INFORMATION: /product= "intron sequence"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2390..2447
OTHER INFORMATION: /product= "intron sequence"
US-08-488-015B-2

Query Match
Best Local Similarity 99.1%; Score 224.8; DB 1; Length 2939;
Pred. No. 1.3e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCATAGTTAAGCCAGCCGCGCAACACCCGCTGAGCGGCC 2410
DB 105 TGCTCTGATGCGCATAGTTAAGCCAGCCGCGCAACACCCGCTGAGCGGCC 164
QY 2411 TGACGGGCTGTGTCTCTCCGCGCATCCGCTTACAGACAAGCTGTGACCGCTCCGGGAGC 2470
DB 165 TGACGGGCTGTGTCTCTCCGCGCATCCGCTTACAGACAAGCTGTGACCGCTCCGGGAGC 224
QY 2471 TGCAATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 2530
DB 225 TGCAATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 284
QY 2531 ATACGCTATTTTATAGTTATGTCATGATATATATGTTCTTAG 2578
DB 285 ATACGCTATTTTATAGTTATGTCATGATATATATGTTCTTAG 332

RESULT 5
US-08-814-412-17
Sequence 17, Application US/08814412
Patent No. 6150141
GENERAL INFORMATION:
APPLICANT: Jarrell Ph.D., Kevin A.
TITLE OF INVENTION: Intron-Mediated Techniques and Reagents
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Choate, Hall & Stewart
STREET: 53 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814,412
FILING DATE: 11-MAR-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jarrell Ph.D., Brenda H.
REGISTRATION NUMBER: 39,223
REFERENCE/DOCKET NUMBER: 0079571-0040
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 248 5000
TELEFAX: 617 248 4000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2939 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "plasmid DNA"
IMMEDIATE SOURCE:
CLONE: BGINV
US-08-814-412-17

Query Match
Best Local Similarity 99.1%; Score 224.8; DB 3; Length 2939;
Pred. No. 1.3e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCATAGTTAAGCCAGCCGCGCAACACCCGCTGAGCGGCC 2410
DB 105 TGCTCTGATGCGCATAGTTAAGCCAGCCGCGCAACACCCGCTGAGCGGCC 164
QY 2411 TGACGGGCTGTGTCTCTCCGCGCATCCGCTTACAGACAAGCTGTGACCGCTCCGGGAGC 2470
DB 165 TGACGGGCTGTGTCTCTCCGCGCATCCGCTTACAGACAAGCTGTGACCGCTCCGGGAGC 224
QY 2471 TGCAATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 2530
DB 225 TGCAATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 284
QY 2531 ATACGCTATTTTATAGTTATGTCATGATATATATGTTCTTAG 2578
DB 285 ATACGCTATTTTATAGTTATGTCATGATATATATGTTCTTAG 332

RESULT 6
US-07-415-307A-1/C
Sequence 1, Application US/07415307A
Patent No. 5344757
GENERAL INFORMATION:
APPLICANT: Holte, Hans Joachim
APPLICANT: Seib, Rudolf
APPLICANT: Schmitz, Gudrun
APPLICANT: Kessler, Hans
APPLICANT: Kessler, Christoph
APPLICANT: Matthes, Ralf
TITLE OF INVENTION: Process for the Detection of Nucleic Acids
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC/DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/415,307A
FILING DATE: 19900109
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP89/0026
FILING DATE: 12-Jan-1989
APPLICATION NUMBER: DE 38 00 642.1

FILING DATE: 12-Jan-1988
APPLICATION NUMBER: DE 38 13 278.8
FILING DATE: 20-Apr-1988
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5344757man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: BOER 798
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-688-9200
TELEFAX: 212-838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3104 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-415-307A-1

Query Match 8.7%; Score 224.8; DB 1; Length 3104;
Best Local Similarity 99.1%; Pred. No. 1,4e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCCTGATGCCGATAGTTAAGCAAGCCCGCAACCCCGCTGACGGCC 2410
DB 2558 TGCTCTGATGCCGATAGTTAAGCAAGCCCGCAACCCCGCTGACGGCC 2499
QY 2411 TGACGGGCTTGTGCTCCCGGATCCGCTTACAGCAAGCTGTGACCGTCCGGAGC 2470
DB 2498 TGACGGGCTTGTGCTCCCGGATCCGCTTACAGCAAGCTGTGACCGTCCGGAGC 2439
QY 2471 TGACGTGTGACAGGTTTTCACCGTCATCACGAAACGCGAGAGCAAGAGGCGCTCGTG 2530
DB 2438 TGACGTGTGACAGGTTTTCACCGTCATCACGAAACGCGAGAGCAAGAGGCGCTCGTG 2379
QY 2531 ATACGCTATTATTATAGTTAATGATGATGATATATATGTTCTTAG 2578
DB 2378 ATACGCTATTATTATAGTTAATGATGATGATATATATGTTCTTAG 2331

RESULT 7

US-08-371-320-1/c
Sequence 1, Application US/08371320
Patent No. 5702888
GENERAL INFORMATION:
APPLICANT: Holtke, Hans Joachim
APPLICANT: Seibl, Rudolf
APPLICANT: Schmitz, Gudrun
APPLICANT: Scholer, Hans
APPLICANT: Kessler, Christoph
TITLE OF INVENTION: Process for the Detection of Nucleic Acids
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC/DOS
SOFTWARE: Nordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/371,320
FILING DATE: 11-Jan-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/415,307
FILING DATE: 09-Jan-1990
APPLICATION NUMBER: PCT/EP89/0026
FILING DATE: 12-Jan-1989

APPLICATION NUMBER: DE 38 00 642.1
FILING DATE: 12-Jan-1988
APPLICATION NUMBER: DE 38 13 278.8
FILING DATE: 20-Apr-1988
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5702888man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: BOER 798
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-688-9200
TELEFAX: 212-838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3104 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-371-320-1

Query Match 8.7%; Score 224.8; DB 1; Length 3104;
Best Local Similarity 99.1%; Pred. No. 1,4e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCCTGATGCCGATAGTTAAGCAAGCCCGCAACCCCGCTGACGGCC 2410
DB 2558 TGCTCTGATGCCGATAGTTAAGCAAGCCCGCAACCCCGCTGACGGCC 2499
QY 2411 TGACGGGCTTGTGCTCCCGGATCCGCTTACAGCAAGCTGTGACCGTCCGGAGC 2470
DB 2498 TGACGGGCTTGTGCTCCCGGATCCGCTTACAGCAAGCTGTGACCGTCCGGAGC 2439
QY 2471 TGACGTGTGACAGGTTTTCACCGTCATCACGAAACGCGAGAGCAAGAGGCGCTCGTG 2530
DB 2438 TGACGTGTGACAGGTTTTCACCGTCATCACGAAACGCGAGAGCAAGAGGCGCTCGTG 2379
QY 2531 ATACGCTATTATTATAGTTAATGATGATGATATATATGTTCTTAG 2578
DB 2378 ATACGCTATTATTATAGTTAATGATGATGATATATATGTTCTTAG 2331

RESULT 8

US-09-027-169-6
Sequence 6, Application US/09027169
Patent No. 6420524
GENERAL INFORMATION:
APPLICANT: CRAIG, NANCY L
TITLE OF INVENTION: GAIN OF FUNCTION MUTATIONS IN
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Brown (Alston & Bird, LLP)
STREET: 3605 Glenwood Ave.
CITY: Raleigh
STATE: NC
COUNTRY: USA
ZIP: 27608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/027,169
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Anne
REGISTRATION NUMBER: 36,463
REFERENCE/DOCKET NUMBER: 5789-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2205
TELEFAX: 919 881 3175
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 3190 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "pRM2 (target plasmid)"
US-09-027-169-6

Query Match 8.7%; Score 224.8; DB 4; Length 3190;
Best Local Similarity 99.1%; Pred. No. 1.4e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGGATGTTAAGCCAGCCCGGACACCCCGGCAACCCCGCTAAGCGGCC 2410
DB 1026 TGCTGTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGGCAACCCCGCTAAGCGGCC 1085
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC 2470
DB 1086 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC 1145
QY 2471 TGATGTGTCAAGAGTTTTCACCGTATCACCGAAACGGCGAGACGAAAGGGCTCTGTG 2530
DB 1146 TGATGTGTCAAGAGTTTTCACCGTATCACCGAAACGGCGAGACGAAAGGGCTCTGTG 1205
QY 2531 ATACGCTATTTTATAGTTAATGATGATATATATATATATATATATATATATATATAT 2578
DB 1206 ATACGCTATTTTATAGTTAATGATGATATATATATATATATATATATATATATAT 1253

RESULT 9
545363-2
PATENT NO. 545363
APPLICANT: RUDOLPH, RAINER; FISCHER, STEPHAN; MATTES, RALF
TITLE OF INVENTION: PROCESS FOR THE ACTIVATION OF T-PA OR
ING AFTER GENETIC EXPRESSION IN PROKARYOTES
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,044
FILING DATE: 02-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 942,370
FILING DATE: 09-SEP-1992
APPLICATION NUMBER: 498,500
FILING DATE: 23-MAR-1990
APPLICATION NUMBER: 76,207
FILING DATE: 23-OCT-1986
SEQ ID NO: 2
LENGTH: 3343
545363-2

Query Match 8.7%; Score 224.8; DB 6; Length 3343;
Best Local Similarity 99.1%; Pred. No. 1.4e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGGATGTTAAGCCAGCCCGGACACCCCGGCAACCCCGCTAAGCGGCC 2410
DB 1179 TGCTGTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGGCAACCCCGCTAAGCGGCC 1238
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC 2470
DB 1239 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC 1298
QY 2471 TGATGTGTCAAGAGTTTTCACCGTATCACCGAAACGGCGAGACGAAAGGGCTCTGTG 2530
DB 1299 TGATGTGTCAAGAGTTTTCACCGTATCACCGAAACGGCGAGACGAAAGGGCTCTGTG 1358
QY 2531 ATACGCTATTTTATAGTTAATGATGATATATATATATATATATATATATATATATAT 2578
DB 1359 ATACGCTATTTTATAGTTAATGATGATATATATATATATATATATATATATATAT 1406

RESULT 10

US-09-470-661A-32/c
Sequence 32, Application US/09470661A
Patent No. 6500662
GENERAL INFORMATION:
APPLICANT: Pfizer Products Inc.
TITLE OF INVENTION: AN INFECTIOUS CDNA CLONE OF NORTH AMERICAN PORCINE
TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME (PRRS) VIRUS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: PC10278A
CURRENT APPLICATION NUMBER: US/09/470,661A
CURRENT FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 32
LENGTH: 3796
TYPE: DNA
ORGANISM: Plasmid
FEATURE:
OTHER INFORMATION: Description of Plasmid: pCMV-MC1
US-09-470-661A-32

Query Match 8.7%; Score 224.8; DB 4; Length 3796;
Best Local Similarity 99.1%; Pred. No. 1.5e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGGATGTTAAGCCAGCCCGGACACCCCGGCAACCCCGCTAAGCGGCC 2410
DB 3564 TGCTGTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGGCAACCCCGCTAAGCGGCC 3505
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC 2470
DB 3504 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC 3445
QY 2471 TGATGTGTCAAGAGTTTTCACCGTATCACCGAAACGGCGAGACGAAAGGGCTCTGTG 2530
DB 3444 TGATGTGTCAAGAGTTTTCACCGTATCACCGAAACGGCGAGACGAAAGGGCTCTGTG 3385
QY 2531 ATACGCTATTTTATAGTTAATGATGATATATATATATATATATATATATATATATAT 2578
DB 3384 ATACGCTATTTTATAGTTAATGATGATATATATATATATATATATATATATATATAT 3337

RESULT 11
US-09-173-053-4/c
Sequence 4, Application US/09173053
Patent No. 6451769
GENERAL INFORMATION:
APPLICANT: HIEBNER, Robert C.
APPLICANT: NORMAN, Jon A.
APPLICANT: LIANG, Xiaowu
APPLICANT: CARNER, Kristin R.
APPLICANT: BARBOUR, Alan G.
APPLICANT: LUXE, Catherine J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ADMINISTERING BORRELLIA DNA
FILE REFERENCE: 454312-2440.1
CURRENT APPLICATION NUMBER: US/09/173,053
CURRENT FILING DATE: 1998-10-15
PRIOR APPLICATION NUMBER: 08/663,998
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 3799
TYPE: DNA
ORGANISM: Borrelia burgdorferi
FEATURE:
NAME/KEY: misc feature
LOCATION: (1118)..(1125)
OTHER INFORMATION: N stands for A or G or C or T
NAME/KEY: misc feature
LOCATION: (1130)..(1137)
OTHER INFORMATION: N stands for A or G or C or T
US-09-173-053-4

Query Match 8.7%; Score 224.8; DB 4; Length 3799;
Best Local Similarity 99.1%; Pred. No. 1.5e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTATGCGCCGATGTTAAGCAGCCCGCAACCCGCTGAGCGGCC 2410
DB 3566 TGCTGTATGCGCCGATGTTAAGCAGCCCGCAACCCGCTGAGCGGCC 3507
QY 2411 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTGACGCTCCCGGAGC 2470
DB 3506 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTGACGCTCCCGGAGC 3447
QY 2471 TGACGTGTGAGAGGTTTTCACCGTCATCAGCGAAGCGGCGAGAGGAGGCTCGTG 2530
DB 3446 TGACGTGTGAGAGGTTTTCACCGTCATCAGCGAAGCGGCGAGAGGAGGCTCGTG 3387
QY 2531 ATACGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578
DB 3386 ATACGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 3339

RESULT 12
US-09-042-353-393/c
Sequence 393, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 393:
SEQUENCE CHARACTERISTICS:
LENGTH: 3819 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-393

Query Match 8.7%; Score 224.8; DB 3; Length 3819;
Best Local Similarity 99.1%; Pred. No. 1.5e-59;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTATGCGCCGATGTTAAGCAGCCCGCAACCCGCTGAGCGGCC 2410
DB 2165 TGCTGTATGCGCCGATGTTAAGCAGCCCGCAACCCGCTGAGCGGCC 2106
QY 2411 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTGACGCTCCCGGAGC 2470
DB 2105 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTGACGCTCCCGGAGC 2046
QY 2471 TGACGTGTGAGAGGTTTTCACCGTCATCAGCGAAGCGGCGAGAGGAGGCTCGTG 2530
DB 2045 TGACGTGTGAGAGGTTTTCACCGTCATCAGCGAAGCGGCGAGAGGAGGCTCGTG 1986
QY 2531 ATACGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578
DB 1985 ATACGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 1938

RESULT 13
US-08-758-417A-243/c
Sequence 243, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 243:
SEQUENCE CHARACTERISTICS:
LENGTH: 3819 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 243:
US-08-758-417A-243
Query Match
Best Local Similarity 99.1%; Score 224.8; DB 4; Length 3819;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TAGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 2410
DB 2165 TGCTCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 2106
QY 2411 TGACGGGCTTGCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCCGGAGGC 2470
DB 2105 TGACGGGCTTGCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCCGGAGGC 2046
QY 2471 TGCAATGTGACAGAGTTTCAACCGTCATACCGAAACCGCGGAGACGAAAGGCGCTG 2530
DB 2045 TGCAATGTGACAGAGTTTCAACCGTCATACCGAAACCGCGGAGACGAAAGGCGCTG 1986
QY 2531 ATAGCCCTATTTTATAGATTATGTCATGATTAATAGTTTCTTAG 2578
DB 1985 ATAGCCCTATTTTATAGATTATGTCATGATTAATAGTTTCTTAG 1938
RESULT 14

US-09-039-982A-21
Sequence 21, Application US/09039982A
Patent No. 6225042
GENERAL INFORMATION:
APPLICANT: Cal, Zelig
APPLICANT: Sprent, Jonathan
APPLICANT: Brunmark, Anders
APPLICANT: Jackson, Michael
APPLICANT: Peterson, Per A
TITLE OF INVENTION: ANTISENSE PRESENTING SYSTEM AND METHODS FOR ACTIVATION OF T-CELL
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Olsson & Hierl, Ltd.
STREET: 20 No. 6225042th Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,982A
FILING DATE: 16-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Olsson, Arne M.
REGISTRATION NUMBER: 30,203
REFERENCE/DOCKET NUMBER: TSRI4710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 580-1180
TELEFAX: (312) 580-1189
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3875 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-039-982A-21
Query Match
Best Local Similarity 99.1%; Score 224.8; DB 3; Length 3875;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TAGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 2410
DB 1482 TGCTCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 1541
QY 2411 TGACGGGCTTGCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCCGGAGGC 2470
DB 1542 TGACGGGCTTGCTGCTCCCGCATCCGCTTACAGACAGCTGTGACCGTCCGGAGGC 1601
QY 2471 TGCAATGTGACAGAGTTTCAACCGTCATACCGAAACCGCGGAGACGAAAGGCGCTG 2530
DB 1602 TGCAATGTGACAGAGTTTCAACCGTCATACCGAAACCGCGGAGACGAAAGGCGCTG 1661
QY 2531 ATAGCCCTATTTTATAGATTATGTCATGATTAATAGTTTCTTAG 2578
DB 1662 ATAGCCCTATTTTATAGATTATGTCATGATTAATAGTTTCTTAG 1709
RESULT 15
US-09-039-641-21
Sequence 21, Application US/09039641
Patent No. 6251627
GENERAL INFORMATION:
APPLICANT: Cal, Zelig
APPLICANT: Sprent, Jonathan

APPLICANT: Brunmark, Anders
 APPLICANT: Jackson, Michael
 APPLICANT: Peterson, Per A
 TITLE OF INVENTION: ANTIGEN PRESENTING SYSTEM AND METHODS FOR
 TITLE OF INVENTION: ACTIVATION OF T-CELLS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Olson & Hierl, Ltd.
 STREET: 20 No. 6251627th Wacker Drive, Suite 3000
 CITY: Chicago
 STATE: Illinois
 COUNTRY: USA
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/039,641
 FILING DATE: 8-MAR-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Olson, Arne M.
 REGISTRATION NUMBER: 30,203
 REFERENCE/DOCKET NUMBER: TSRI4710
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 580-1180
 TELEFAX: (312) 580-1189
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3875 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-09-039-641-21

Query Match 8.7%; Score 224.8; DB 3; Length 3875;
 Best Local Similarity 99.1%; Freq. No. 1.6e-59;
 Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2351 TACGCTGATGCCGATGTTAAGCCAGCCCGCAACACCCGCTGACGGCCC 2410
 Db 1482 TGCTGATGAGCCGATGTTAAGCCAGCCCGCAACACCCGCTGACGGCCC 1541
 QY 2411 TGACGAGCTTGTGCTGCCCGCATCCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 2470
 Db 1542 TGACGAGCTTGTGCTGCCCGCATCCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 1601
 QY 2471 TGATGATGACAGAGTTTACCGGATGACCGGAAAGCGGAGAGCAAGGGCTCGTG 2530
 Db 1602 TGATGATGACAGAGTTTACCGGATGACCGGAAAGCGGAGAGCAAGGGCTCGTG 1661
 QY 2531 ATAGGCTATTTTATAGGTTAATGTCATGATATATATGTTCTTAG 2578
 Db 1662 ATAGGCTATTTTATAGGTTAATGTCATGATATATATGTTCTTAG 1709

Search completed: April 30, 2004, 18:42:53
 Job time: 307.986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 30, 2004, 15:29:25 ; Search time 1143.74 Seconds
(without alignments)
10195.950 Million cell updates/sec

Title: US-10-603-260-1

Perfect score: 2578
Sequence: 1 agcttcacattatcagcga.....tgataataatggtttcttag 2578

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

| | |
|-----|---|
| 1: | /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:* |
| 2: | /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:* |
| 3: | /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:* |
| 4: | /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:* |
| 5: | /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:* |
| 6: | /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUBCOMB.seq:* |
| 7: | /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:* |
| 8: | /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:* |
| 9: | /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:* |
| 10: | /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:* |
| 11: | /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:* |
| 12: | /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:* |
| 13: | /cgn2_6/ptodata/2/pubpna/US09_NEW_PUBCOMB.seq:* |
| 14: | /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:* |
| 15: | /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:* |
| 16: | /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:* |
| 17: | /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:* |
| 18: | /cgn2_6/ptodata/2/pubpna/US10_NEW_PUBCOMB.seq:* |
| 19: | /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | ID | Description |
|------------|-------|---------------|--------|----------------------|--------------------|
| 1 | 2578 | 100.0 | 2578 | US-10-603-260-1 | Sequence 1, Appl1 |
| 2 | 798 | 31.0 | 798 | US-10-603-260-2 | Sequence 2, Appl1 |
| 3 | 555 | 21.5 | 555 | US-10-603-260-4 | Sequence 4, Appl1 |
| 4 | 408.2 | 15.8 | 717 | US-10-282-122A-41398 | Sequence 41398, A |
| 5 | 249.2 | 9.7 | 1308 | US-10-282-122A-41399 | Sequence 41399, A |
| 6 | 224.8 | 8.7 | 878 | US-09-764-847-1872 | Sequence 1872, Ap |
| 7 | 224.8 | 8.7 | 878 | US-09-764-891-10029 | Sequence 10029, A |
| 8 | 224.8 | 8.7 | 878 | US-09-764-891-10033 | Sequence 10033, A |
| 9 | 224.8 | 8.7 | 878 | US-09-973-278-674 | Sequence 674, App |
| 10 | 224.8 | 8.7 | 878 | US-10-092-154-1872 | Sequence 1872, Ap |
| 11 | 224.8 | 8.7 | 1073 | US-09-788-987-6 | Sequence 6, Appl1 |
| 12 | 224.8 | 8.7 | 2297 | US-09-891-865A-11 | Sequence 11, Appl1 |
| 13 | 224.8 | 8.7 | 2462 | US-10-365-862-5 | Sequence 5, Appl1 |
| 14 | 224.8 | 8.7 | 2762 | US-10-244-142A-2 | Sequence 2, Appl1 |

| | | | | | | | |
|---|----|-------|-----|------|----|-------------------|--------------------|
| C | 15 | 224.8 | 8.7 | 2803 | 15 | US-10-244-142A-1 | Sequence 1, Appl1 |
| | 16 | 224.8 | 8.7 | 2880 | 15 | US-10-128-590-62 | Sequence 62, Appl1 |
| | 17 | 224.8 | 8.7 | 2880 | 15 | US-10-128-587A-62 | Sequence 62, Appl1 |
| | 18 | 224.8 | 8.7 | 2880 | 15 | US-10-128-578B-62 | Sequence 62, Appl1 |
| | 19 | 224.8 | 8.7 | 3031 | 10 | US-09-891-865A-12 | Sequence 9, Appl1 |
| | 20 | 224.8 | 8.7 | 3067 | 15 | US-10-128-590-9 | Sequence 12, Appl1 |
| | 21 | 224.8 | 8.7 | 3067 | 15 | US-10-128-587A-9 | Sequence 9, Appl1 |
| | 22 | 224.8 | 8.7 | 3067 | 16 | US-10-128-578B-9 | Sequence 9, Appl1 |
| | 23 | 224.8 | 8.7 | 3128 | 10 | US-09-891-865A-13 | Sequence 13, Appl1 |
| | 24 | 224.8 | 8.7 | 3126 | 14 | US-10-024-809-6 | Sequence 6, Appl1 |
| | 25 | 224.8 | 8.7 | 3216 | 15 | US-10-057-108-1 | Sequence 1, Appl1 |
| | 26 | 224.8 | 8.7 | 3309 | 13 | US-10-343-303-7 | Sequence 7, Appl1 |
| | 27 | 224.8 | 8.7 | 3383 | 10 | US-09-891-865A-3 | Sequence 3, Appl1 |
| | 28 | 224.8 | 8.7 | 3444 | 10 | US-09-891-865A-1 | Sequence 1, Appl1 |
| | 29 | 224.8 | 8.7 | 3561 | 16 | US-10-382-361-2 | Sequence 2, Appl1 |
| | 30 | 224.8 | 8.7 | 3730 | 15 | US-10-128-590-22 | Sequence 22, Appl1 |
| | 31 | 224.8 | 8.7 | 3730 | 15 | US-10-128-587A-22 | Sequence 22, Appl1 |
| | 32 | 224.8 | 8.7 | 3730 | 16 | US-10-128-578B-22 | Sequence 22, Appl1 |
| | 33 | 224.8 | 8.7 | 3753 | 15 | US-10-258-482-3 | Sequence 3, Appl1 |
| | 34 | 224.8 | 8.7 | 3755 | 15 | US-10-258-482-4 | Sequence 4, Appl1 |
| | 35 | 224.8 | 8.7 | 3796 | 14 | US-10-127-391-32 | Sequence 32, Appl1 |
| | 36 | 224.8 | 8.7 | 3796 | 15 | US-10-241-332-32 | Sequence 32, Appl1 |
| | 37 | 224.8 | 8.7 | 3858 | 16 | US-10-014-099F-14 | Sequence 14, Appl1 |
| | 38 | 224.8 | 8.7 | 3875 | 15 | US-10-105-200A-21 | Sequence 21, Appl1 |
| | 39 | 224.8 | 8.7 | 3875 | 15 | US-10-105-504A-21 | Sequence 21, Appl1 |
| | 40 | 224.8 | 8.7 | 3875 | 15 | US-10-105-678A-21 | Sequence 21, Appl1 |
| | 41 | 224.8 | 8.7 | 3875 | 15 | US-10-266-463A-21 | Sequence 21, Appl1 |
| | 42 | 224.8 | 8.7 | 3878 | 15 | US-10-105-200A-27 | Sequence 27, Appl1 |
| | 43 | 224.8 | 8.7 | 3878 | 15 | US-10-105-678A-27 | Sequence 27, Appl1 |
| | 44 | 224.8 | 8.7 | 3878 | 15 | US-10-105-678A-27 | Sequence 27, Appl1 |
| | 45 | 224.8 | 8.7 | 3878 | 15 | US-10-266-463A-27 | Sequence 27, Appl1 |

ALIGNMENTS

RESULT 1
US-10-603-260-1
; Sequence 1, Application US/10603260
; Publication No. US20040009570A1
; GENERAL INFORMATION:
; APPLICANT: CJ Corporation
; TITLE OF INVENTION: An alkaline lipase from *Vibrio metschnikovii* RH530 and a
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/603,260
; CURRENT FILING DATE: 2003-06-24
; NUMBER OF SEQ. ID NOS: 5
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 1
; LENGTH: 2578
; TYPE: DNA
; ORGANISM: *Vibrio metschnikovii* RH530
US-10-603-260-1

Query Match 100.0%; Score 2578; DB 16; Length 2578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | AGCTTCACATTATCAGCAATACCTGATCGTACCTGCGGCGGCACTTGCCAGTG | 60 |
| DB | 1 | AGCTTCACATTATCAGCAATACCTGATCGTACCTGCGGCGGCACTTGCCAGTG | 60 |
| QY | 61 | CGCGCGGCTACGTAAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATG | 120 |
| DB | 61 | CGCGCGGCTACGTAAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATG | 120 |
| QY | 121 | TGCGCGGCTACGTAAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATG | 180 |
| DB | 121 | TGCGCGGCTACGTAAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATG | 180 |
| QY | 181 | GATTTCATTTGGCGGCTGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATG | 240 |

Db 181 GATTTCATTGGCGATCGGGTTGAAAATGGAATAGCGTAATGACTGTAAATAAGTACG 240
QY 241 ATTCAATGAGTGATGCTGCTCTAATAATAAACATGTGGGATCGGAAAGCGCAATGA 300
Db 241 ATTCAATGAGTGATGCTGCTCTAATAATAAACATGTGGGATCGGAAAGCGCAATGA 300
QY 301 AGGCACTGATGTAATTTCTTGGGCTACTCTCTTAATAATGCTTAATCTTCTACT 360
Db 301 AGGCACTGATGTAATTTCTTGGGCTACTCTCTTAATAATGCTTAATCTTCTACT 360
QY 361 GCGCAATCAATTCATAGGCGACCTCAACATCGGATCAACGATGACCGGCTGCAT 420
Db 361 GCGCAATCAATTCATAGGCGACCTCAACATCGGATCAACGATGACCGGCTGCAT 420
QY 421 CATATTGATTTATAGGCAATCAAGATCAATTAACCTGGATTAAGAGGAAATAG 480
Db 421 CATATTGATTTATAGGCAATCAAGATCAATTAACCTGGATTAAGAGGAAATAG 480
QY 481 GCGATGCGCTGATGATGCGATCGCTTGGTGGTTGATTCAGTCATTAAGTATGTA 540
Db 481 GCGATGCGCTGATGATGCGATCGCTTGGTGGTTGATTCAGTCATTAAGTATGTA 540
QY 541 CAACGTGTTCTAATCTTGGGCGAACATTAAGTACCTTGTAAAGTTTGTCAACTTTGT 600
Db 541 CAACGTGTTCTAATCTTGGGCGAACATTAAGTACCTTGTAAAGTTTGTCAACTTTGT 600
QY 601 GACAAACCTAGTACGCTTATTTGGCCTTAATTAATTAAGATTAAGAGGATAGGA 660
Db 601 GACAAACCTAGTACGCTTATTTGGCCTTAATTAATTAAGATTAAGAGGATAGGA 660
QY 661 CGTAGTCATTAACAATTAACAGTACTCTGTATCTGATTAATGTTGTCAAAAGTCT 720
Db 661 CGTAGTCATTAACAATTAACAGTACTCTGTATCTGATTAATGTTGTCAAAAGTCT 720
QY 721 TATTACATTTGACCATCATCATGCACTTAACCTTAATAAGCCGCTGTTTATAGGAA 780
Db 721 TATTACATTTGACCATCATCATGCACTTAACCTTAATAAGCCGCTGTTTATAGGAA 780
QY 781 GGCATATGATGTCTACTATCATGATGATTTGTCTGCTTGGCGCAATCTATCCAG 840
Db 781 GGCATATGATGTCTACTATCATGATGATTTGTCTGCTTGGCGCAATCTATCCAG 840
QY 841 GTTTTACTGGTAAAGCTCTTAATCCAAATCGGCAATGTGTAAATGGGCAATGCT 900
Db 841 GTTTTACTGGTAAAGCTCTTAATCCAAATCGGCAATGTGTAAATGGGCAATGCT 900
QY 901 GGGGGGTAAGTATGACGAAGATGACCGCTCATGTGAGGAACCTTCATGAGAT 960
Db 901 GGGGGGTAAGTATGACGAAGATGACCGCTCATGTGAGGAACCTTCATGAGAT 960
QY 961 TTTGATGACGAGACGACGATTTGTGGCAAAAAGTCATATCTTATCTTAATTTATC 1020
Db 961 TTTGATGACGAGACGACGATTTGTGGCAAAAAGTCATATCTTATCTTAATTTATC 1020
QY 1021 AGGATTCGGTGTGATGAGCAACCCCAACGGGATCCGAATGCTTGGAGTTCAGTAT 1080
Db 1021 AGGATTCGGTGTGATGAGCAACCCCAACGGGATCCGAATGCTTGGAGTTCAGTAT 1080
QY 1081 TCCCATTAACGCTTATTAACCCCGTGAATGTCAACCAATAGAAATTTGGTATCAG 1140
Db 1081 TCCCATTAACGCTTATTAACCCCGTGAATGTCAACCAATAGAAATTTGGTATCAG 1140
QY 1141 CCGGAGGCGCTTAATGTTGATCTTCAATCTTACTCAAGAGAAATGCGCGTGGCT 1200
Db 1141 CCGGAGGCGCTTAATGTTGATCTTCAATCTTACTCAAGAGAAATGCGCGTGGCT 1200
QY 1201 TTTGATCATGTGCGCAAAATTCAGACGATGCGCAAAATTCAGCGCTCGGTTGAATAC 1260
Db 1201 TTTGATCATGTGCGCAAAATTCAGACGATGCGCAAAATTCAGCGCTCGGTTGAATAC 1260
QY 1261 ACATCGTGTACTATTTTCAATTAAGAAAGAGTTTATGTCGCGATTTATGATGCC 1320
Db 1261 ACATCGTGTACTATTTTCAATTAAGAAAGAGTTTATGTCGCGATTTATGATGCC 1320

QY 1321 TAGCCAAATTTGGCGTAGATTAAATCGATGACATTAACCGGCTTGATCAATACC 1380
Db 1321 TAGCCAAATTTGGCGTAGATTAAATCGATGACATTAACCGGCTTGATCAATACC 1380
QY 1381 GGGGTGATGCTGATCAATTAATAATGCGGCAATCTTGTAAAGCAAGAGCAACCA 1440
Db 1381 GGGGTGATGCTGATCAATTAATAATGCGGCAATCTTGTAAAGCAAGAGCAACCA 1440
QY 1441 GCGACGTTATCGCTTTCAGTACATGAGTCACTTAATTTCAAACTGTTTACGAGGT 1500
Db 1441 GCGACGTTATCGCTTTCAGTACATGAGTCACTTAATTTCAAACTGTTTACGAGGT 1500
QY 1501 TAACTGTTGAAAATCGGTACAGTAGATGATGTCATTTGATGATAGTAGGAAGA 1560
Db 1501 TAACTGTTGAAAATCGGTACAGTAGATGATGTCATTTGATGATAGTAGGAAGA 1560
QY 1561 ATGAGATTAATCTTGTTCATGACCTATATGATGATGCTTGGTAAATGATCCGCTTATG 1620
Db 1561 ATGAGATTAATCTTGTTCATGACCTATATGATGATGCTTGGTAAATGATCCGCTTATG 1620
QY 1621 CATGCTGCAATTAATGGGTTATCGTATCAAACTTAAGTAACTCACTGCTATC 1680
Db 1621 CATGCTGCAATTAATGGGTTATCGTATCAAACTTAAGTAACTCACTGCTATC 1680
QY 1681 GATGATGAGGCAATTTTGGCGGCTTGAACGATCGCTCACTGATGCTGCTATGCT 1740
Db 1681 GATGATGAGGCAATTTTGGCGGCTTGAACGATCGCTCACTGATGCTGCTATGCT 1740
QY 1741 TTATGCGGACACAGTTTGGCGGATGATGATCAACGTTATCTAAGTATGCGGCAACG 1800
Db 1741 TTATGCGGACACAGTTTGGCGGATGATGATCAACGTTATCTAAGTATGCGGCAACG 1800
QY 1801 TCCGTGAAACCCCTCCATGCTGCTGCGATGCGGCTCACTTGAAGAGGCTTCAT 1860
Db 1801 TCCGTGAAACCCCTCCATGCTGCTGCGATGCGGCTCACTTGAAGAGGCTTCAT 1860
QY 1861 GTCAATTAATTTGACATTAATGATTTAGGGGTGACCTAGATTAATCAGCAGAAATTTGG 1920
Db 1861 GTCAATTAATTTGACATTAATGATTTAGGGGTGACCTAGATTAATCAGCAGAAATTTGG 1920
QY 1921 TTTAAAGAACACGACGACGATCCGCTATCCCAAAAATCAGGACATTTGACGAGAAC 1980
Db 1921 TTTAAAGAACACGACGACGATCCGCTATCCCAAAAATCAGGACATTTGACGAGAAC 1980
QY 1981 ATACCTTTAGGGCTGCGCAGGCTTTTACTGCGGATCCACTGACCTCGATGTACCGTC 2040
Db 1981 ATACCTTTAGGGCTGCGCAGGCTTTTACTGCGGATCCACTGACCTCGATGTACCGTC 2040
QY 2041 ACAGTAGAAGAACCAAAATAGCTGCGATGACAGATATATCGGATATCACCACTTCA 2100
Db 2041 ACAGTAGAAGAACCAAAATAGCTGCGATGACAGATATATCGGATATCACCACTTCA 2100
QY 2101 TAGAGAAATGCTTTAATGATTCGTTGCGAGCAATGACCACTTTCTTGATAGA 2160
Db 2101 TAGAGAAATGCTTTAATGATTCGTTGCGAGCAATGACCACTTTCTTGATAGA 2160
QY 2161 CCGCTTCCGCGCTAAAGCCGTTTAACTTCAAGTATGATGTAATCTTGATCAACCGA 2220
Db 2161 CCGCTTCCGCGCTAAAGCCGTTTAACTTCAAGTATGATGTAATCTTGATCAACCGA 2220
QY 2221 TGGTGAATTAAGAAATATCCCACTTCAATCAAGATTAAGAGGTTGCAATGAGAGCTT 2280
Db 2221 TGGTGAATTAAGAAATATCCCACTTCAATCAAGATTAAGAGGTTGCAATGAGAGCTT 2280
QY 2281 TCCCATGCAATTAACATTCGCGCATTTACGCTGCGCTTTAATCTAAATTTGGCAATG 2340
Db 2281 TCCCATGCAATTAACATTCGCGCATTTACGCTGCGCTTTAATCTAAATTTGGCAATG 2340
QY 2341 TCTGCCGGAATAGCTGATGCGGATAGTTAAGCAGCCCGGACACCCGCAACCCGC 2400
Db 2341 TCTGCCGGAATAGCTGATGCGGATAGTTAAGCAGCCCGGACACCCGCAACCCGC 2400

QY 2401 TGACGCGCCCTGACGCGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGACGTCGACCT 2460
DB 2401 TACGCGCCCTGACGCGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGACGTCGACCT 2460
QY 2461 CTCGCGGAGCTGCATGTGTCAAGAGTTTTCACCGTCATCACGGAACGCGGAGACGAA 2520
DB 2461 CTCGCGGAGCTGCATGTGTCAAGAGTTTTCACCGTCATCACGGAACGCGGAGACGAA 2520
QY 2521 GCGGCTCGGATGACCGCTATTTTATAGTTATGTCAATGATATATATGTTTCTTAA 2578
DB 2521 GCGGCTCGGATGACCGCTATTTTATAGTTATGTCAATGATATATATGTTTCTTAA 2578

RESULT 2
US-10-603-260-2
Sequence 2, Application US/10603260
Publication No. US20040009570A1
GENERAL INFORMATION:
APPLICANT: CJ Corporation
TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/603,260
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Koparentin 1.71
SEQ ID NO 2
LENGTH: 798
TYPE: DNA
ORGANISM: *Vibrio metchnikovii* RH530
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(798)
OTHER INFORMATION: val11 gene
US-10-603-260-2

Query Match 31.0%; Score 798; DB 16; Length 798;
Best Local Similarity 100.0%; Pred. No. 8e-232;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 ATGTTTGCACAAAGTCTTATTTACATTTGACATGACATGACATGACATGACATGACATGAC 762
DB 1 ATGTTTGCACAAAGTCTTATTTACATTTGACATGACATGACATGACATGACATGACATGAC 60
QY 763 CCGTTGTTTATAGGAGAACCATTAATGATGTCATGATGATGATGATGATGATGATGATGATG 822
DB 61 CCGTTGTTTATAGGAGAACCATTAATGATGTCATGATGATGATGATGATGATGATGATGATG 120
QY 823 GCGCGGAATCTATTCAGAGTTTACTGTTGAAACGCTCTAATCCAAATCGGCGAATGTT 882
DB 121 GCGCGGAATCTATTCAGAGTTTACTGTTGAAACGCTCTAATCCAAATCGGCGAATGTT 180
QY 883 GGTAAATGGGCAATGCTCGCGGAGATAGTATGATGACGAAGATGACGCGCATGTTGGA 942
DB 181 GGTAAATGGGCAATGCTCGCGGAGATAGTATGATGACGAAGATGACGCGCATGTTGGA 240
QY 943 GAACCTGTGATGAGATTTTGTATGACAGGACGACGATTTGTGCGCAAAAAGTCCAT 1002
DB 241 GAACCTGTGATGAGATTTTGTATGACAGGACGACGATTTGTGCGCAAAAAGTCCAT 300
QY 1003 ACTATCTATTTTATGAGGATCGGCTGTTGATGAGCAACCCCAACGCGATCCGAAT 1062
DB 301 ACTATCTATTTTATGAGGATCGGCTGTTGATGAGCAACCCCAACGCGATCCGAAT 360
QY 1063 GGTGAGTGTCAATTTTCCATTAAGCTTTATTAACCGGTGAAATGTCAAACAAATA 1122
DB 361 GGTGAGTGTCAATTTTCCATTAAGCTTTATTAACCGGTGAAATGTCAAACAAATA 420
QY 1123 GAATATTTTGTATGACCGCGAGCGGCTAATGTTGATGATCTTCACTTAACTTAA 1182
DB 421 GAATATTTTGTATGACCGCGAGCGGCTAATGTTGATGATCTTCACTTAACTTAA 480
QY 1183 GAAGAAATGCGCTGTTTGTATCATGTGCGCAAAATGACATGCGTGGCAAAAATTA 1242

DB 481 GAAAGAAATGCGCTGCTTTTGTATCATGTGCGCAAAATGACATGCGTGGCAAAAATTA 540
QY 1243 CCGCTGCGGTTGAATACACATCCGTGTACTATTTTCAATTTGAAAAAGGTTTAA 1302
DB 541 CCGCTGCGGTTGAATACACATCCGTGTACTATTTTCAATTTGAAAAAGGTTTAA 600
QY 1303 GCGGATTTATTTATGATGCTTACGCGCAAAATTTGCGGTGAAATTAATGACATGACATTA 1362
DB 601 GCGGATTTATTTATGATGCTTACGCGCAAAATTTGCGGTGAAATTAATGACATGACATTA 660
QY 1363 CCGCTGTTATCAATACCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1422
DB 661 CCGCTGTTATCAATACCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 1423 GGCAGAGAGCCAAACAGCCACCGTTTATGCTTTCAGATGATGATGATGATGATGATGAT 1482
DB 721 GGCAGAGAGCCAAACAGCCACCGTTTATGCTTTCAGATGATGATGATGATGATGATGAT 780
QY 1483 CAACCTGTTTACGAGT 1500
DB 781 CAACCTGTTTACGAGT 798

RESULT 3
US-10-603-260-4
Sequence 4, Application US/10603260
Publication No. US20040009570A1
GENERAL INFORMATION:
APPLICANT: CJ Corporation
TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/603,260
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Koparentin 1.71
SEQ ID NO 4
LENGTH: 555
TYPE: DNA
ORGANISM: *Vibrio metchnikovii* RH530
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(555)
OTHER INFORMATION: val12 gene
US-10-603-260-4

Query Match 21.5%; Score 555; DB 16; Length 555;
Best Local Similarity 100.0%; Pred. No. 7.5e-158;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 ATGCAATTAATCTTGTTCATGACCTATATGATGATGATGATGATGATGATGATGATGATG 1620
DB 1 ATGCAATTAATCTTGTTCATGACCTATATGATGATGATGATGATGATGATGATGATGATG 60
QY 1621 CATGCTGCAATTAATTTGGTTATGCTCAACCAATTAATGATCACTCACTGCTATC 1680
DB 61 CATGCTGCAATTAATTTGGTTATGCTCAACCAATTAATGATCACTCACTGCTATC 120
QY 1681 GATGATGAGGCAATTTTTCGCGCTTACGCAATGCTCACTCACTCACTGCTGCTATGCT 1740
DB 121 GATGATGAGGCAATTTTTCGCGCTTACGCAATGCTCACTCACTCACTGCTGCTATGCT 180
QY 1741 TTATGTCGACACAGTTTGGCGGATGATGATCAACGTTATCTAAGAAATGCGCGCACCG 1800
DB 181 TTATGTCGACACAGTTTGGCGGATGATGATCAACGTTATCTAAGAAATGCGCGCACCG 240
QY 1801 TCGTGTGAAACCTCTCCCATGTGTCGCAATGCGCTCACTTTGCAAGAGGCTTCAAT 1860
DB 241 TCGTGTGAAACCTCTCCCATGTGTCGCAATGCGCTCACTTTGCAAGAGGCTTCAAT 300
QY 1861 GTCAATTAATTAATGACATTAATGTTTGAAGGTTGCACTAGGTAATTCAGCAAAATTTGG 1920

Db 301 GTCAATTAATTCAGATTTAGCTTTAGGGGTGGCACTAGATTAATTCAGAGAAATTTGGG 360
QY 1921 TTAATAACACACGACGACGATCCGCTATCCAAAAATGAGGAGATTCAGAGAAAG 1980
Db 361 TTAATAACACACGACGACGATCCGCTATCCAAAAATGAGGAGATTCAGAGAAAG 420
QY 1981 ATACCTTTGGGGTGGGCGACCTTTTACGCGCGATCCAGTCCGATGCTACCGTC 2040
Db 421 ATACCTTTGGGGTGGGCGACCTTTTACGCGCGATCCAGTCCGATGCTACCGTC 480
QY 2041 ACAGTAGAAGAAACCAAAATAGCTGCGATGACGATCATATGCGATATCCACCACTTCA 2100
Db 481 ACAGTAGAAGAAACCAAAATAGCTGCGATGACGATCATATGCGATATCCACCACTTCA 540
QY 2101 TACGAGAAATGCTTT 2115
Db 541 TACGAGAAATGCTTT 555

RESULT 4

US-10-282-122A-41398
Sequence 41398, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 41398
LENGTH: 717
TYPE: DNA
ORGANISM: Vibrio cholerae
US-10-282-122A-41398

Query Match 15.8%; Score 408.2; DB 13; Length 717;
Best Local Similarity 73.1%; Pred. No. 5.2e-113;
Matches 524; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 787 ATGATTTGCTACTATCCATATGATTTGTCTGCGCTTTGCGCCGAATCTATCCAGTTTAA 846
Db 1 ATGATTTGCTACTATCCATATGATTTGTCTGCGCTTTGCGCCGAATCTATCCAGTTTAA 846
QY 847 CTGGTGAAGAGCTCTTAATCCAAATCGGCCGAGATTTGTGTAAATGGGCAATTCCTGGCGGG 906
Db 61 TTGGTCAAGAGCGAGCAATCCAAATCGGCCGAGATTTGTGTAAATGGGCAATTCCTGGCGGG 906
QY 907 ATAGTATAGAGAGATATGATGACCGCTCATGTGTGAGAACTGTGATGATGATTTTGAAT 966
Db 121 TGGGTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 967 GAGGAGAGAGAGATTTGTGTGCGCAAAAGTCCATATGATGATGATGATGATGATGATGAT 1026
Db 181 TCGGCGGTGCGGAGATCTGCGCCGAGAAATCATATCTTACCTTCACTTCACTGAT 240
QY 1027 CCGCTGTTGATGAGCAACCCCAAGCGATCCGAATGTTGATGATGATGATGATGATGAT 1086
Db 241 CTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 1087 TAGCTTTATTAACCGTGGAGATGTCACAAATGAAAGATTTGTGATGATGATGATGAT 1146
Db 301 TAGGCTGCTCAATCAATTCACAGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 1147 CGGCTAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1206
Db 361 CGGCTAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1266
QY 1207 CATGTGCGCAATTCAGCATGCGTGGCAAAATTAAGCGTGGTGGTGAATACATCC 1266
Db 421 CATGTGCGCAATTCAGCATGCGTGGCAAAATTAAGCGTGGTGGTGAATACATCC 1266
QY 1267 GTGTACTATTTTCAATAGAAAAGATTTTGTGGCGGATTTTATGATGATGATGATGAT 1326
Db 481 GTGTACTATTTTCAATAGAAAAGATTTTGTGGCGGATTTTATGATGATGATGATGAT 1386
QY 1327 AAATTTGGCGTCAAGTAAATGCGATGACCATTAAGCGCGCTTGTATCAATACCGGGGTG 1386
Db 541 AAATTTGGCGTCAAGTAAATGCGATGACCATTAAGCGCGCTTGTATCAATACCGGGGTG 1446
QY 1387 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1446
Db 601 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1503
QY 1447 GTTTATGCTCTTCCAGTCAAGTCAATATTTTCAACCTGTTTACGAGTTAA 1503
Db 661 GTTTATGCTCTTCCAGTCAAGTCAATATTTTCAACCTGTTTACGAGTTAA 1503

RESULT 5

US-10-282-122A-41399/c
Sequence 41399, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848

Prior Filing Date: 2000-05-23
Prior Application Number: 60/207,727
Prior Filing Date: 2000-05-26
Prior Application Number: 60/230,335
Prior Filing Date: 2000-09-06
Prior Application Number: 60/230,347
Prior Filing Date: 2000-09-09
Prior Application Number: 60/242,578
Prior Filing Date: 2000-10-23
Prior Application Number: 60/253,625
Prior Filing Date: 2000-11-27
Prior Application Number: 60/257,931
Prior Filing Date: 2000-12-22
Prior Application Number: 60/267,636
Prior Filing Date: 2001-02-09
Prior Application Number: 60/269,308
Prior Filing Date: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
Number of SEQ ID NOS: 78614
Software: Patent version 3.1
SEQ ID NO 41399
Length: 1308
Type: DNA
Organism: Vibrio cholerae
US-10-282-122A-41399

Query Match
Best Local Similarity 9.7%; Score 249.2; DB 13; Length 1308;
Pred. No. 2.4e-64;
Matches 332; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

23 ACTTGATCGTGAACCTGCGGCGGCACTTGTGCGGCGGCGGCTAGTCTCAAG 82
470 ACCTTCAAGGTAAATCAGAGGCACTCAGGCCAAGCTGACGCTGTGATCTCAAT 411
83 ATTAAAGCCATGACTAGCGTTCTATTAATGATGATGCTCCGACGATCTGAATGCG 142
410 ACGATGCGCATGACTAGGTTCTATCAAAATGATGCTCAAGCACTGCAAGAACTG 351
143 ATACGAGCGTGGGCTTGGCCCTTGTGAGGATCCGATTCATTTGCGGATCGGAT 202
350 ATCCGAGTGTCTGTTTCCGCTTGTGCAATCCCATTTCTTACTTCTTGAAGC 291
203 TGAATAATGAAATAGCGTATGATCTGTAAAAAGTACGATCAATGAGTGCATGCTGC 262
290 ACAAATGAAATAGCAACGATGTCAGAAAGTGTGCTTGAATGATGCGCATGCTGA 231
263 TCTAATAAACAATGTGCGCATCCGAAAAGCGCATGAAAGCACTGATGATTTCTTGG 322
230 GTCAAAATTAATGAAATATCGGATCCGAAAAGCGCAATGCGCAAGATGAGCTATCTTGA 171
323 CGTACTCTCTATTAATGCTAATGTCTCATCACTGCGACATCAATTCATAGGCG 382
170 CGGATGCGATCAAGAGTCCGCTGCGCATCTCTTCACTGCGATCAATTCATAGGCG 111
383 ACTTCAATCCGATACACGATGAACGCGCTGATATTTGATTTTATAGGATCA 442
110 ACCGACATCCGATAAAAATGATGATGCTGATCATTTATTTATAGCATCA 51
443 AGATCAATTAATCTGCGATTAAGAGGAAATAGGCGATGCTGCT 492
50 AGATCGAGCGGCTATGAAATGATGTGGGAAAGAAACGCGGTGATTCAT 1

RESULT 6
US-09-764-847-1872
Sequence 1872, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper
Number of SEQ ID NOS: 2003
Software: Patentin Ver. 2.0
SEQ ID NO 1872
Length: 878
Type: DNA
Organism: Homo sapiens
Feature:
Name/Key: SITE
Location: (741)
OTHER INFORMATION: n equals a,t,c, or c
US-09-764-847-1872

Query Match
Best Local Similarity 8.7%; Score 224.8; DB 9; Length 878;
Pred. No. 5e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2351 TAGCGTATGCGGATGATTAAGCGAGCCCGACACCCCGCAACCCGCTGACGCGCC 2410
202 TGCTTATGCGGATGATTAAGCGAGCCCGACACCCCGCAACCCGCTGACGCGCC 261
2411 TGAAGGCTTGTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCTCGGAGC 2470
262 TGAAGGCTTGTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCTCGGAGC 321
2471 TGCAATGTGAGAGTTTTCACCGTATCAACGAAAGCGCGAGAGAGGCGCTTGTG 2530
322 TGCAATGTGAGAGTTTTCACCGTATCAACGAAAGCGCGAGAGAGGCGCTTGTG 381
2531 ATACGCTATTTTATAGTTAATGTCATGATATATAGTTTCTTGA 2578
382 ATACGCTATTTTATAGTTAATGTCATGATATATAGTTTCTTGA 429

RESULT 7
US-09-764-891-10029
Sequence 10029, Application US/09764891
Publication No. US2003007808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
Number of SEQ ID NOS: 10231
Software: Patentin Ver. 2.0
SEQ ID NO 10029
Length: 878
Type: DNA
Organism: Homo sapiens
Feature:
Name/Key: SITE
Location: (741)
OTHER INFORMATION: n equals a,t,c, or c
US-09-764-891-10029

Query Match
Best Local Similarity 8.7%; Score 224.8; DB 10; Length 878;
Pred. No. 5e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2351 TAGCGTATGCGGATGATTAAGCGAGCCCGACACCCCGCAACCCGCTGACGCGCC 2410
202 TGCTTATGCGGATGATTAAGCGAGCCCGACACCCCGCAACCCGCTGACGCGCC 261
2411 TGAAGGCTTGTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCTCGGAGC 2470
262 TGAAGGCTTGTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCTCGGAGC 321
2471 TGCAATGTGAGAGTTTTCACCGTATCAACGAAAGCGCGAGAGAGGCGCTTGTG 2530
322 TGCAATGTGAGAGTTTTCACCGTATCAACGAAAGCGCGAGAGAGGCGCTTGTG 381

QY 2531 ATACGGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578
DB 382 ATACGGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 429

RESULT 8
US-09-764-891-10033
Sequence 10033, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10033
LENGTH: 878
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (741)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-10033

Query Match 8.7%; Score 224.8; DB 10; Length 878;
Best Local Similarity 99.1%; Pred. No. 5e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGCATAGTTAAGCCAGCCCGGCAACCCGCTGAGCGGCC 2410
DB 202 TGCTTGATGCGCATAGTTAAGCCAGCCCGGCAACCCGCTGAGCGGCC 261

QY 2411 TGAAGGAGCTTGTGCTCCCGCATCCGTTACAGCAAGCTGTACCGTCTCCGGAGC 2470
DB 262 TGAAGGAGCTTGTGCTCCCGCATCCGTTACAGCAAGCTGTACCGTCTCCGGAGC 321

QY 2471 TGATGTGTACAGAGGTTTACCGTATCACCGAAGCGGCGAGAGAAAGGCGCTCGTG 2530
DB 322 TGATGTGTACAGAGGTTTACCGTATCACCGAAGCGGCGAGAGAAAGGCGCTCGTG 381

QY 2531 ATACGGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578
DB 382 ATACGGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 429

RESULT 9
US-09-973-278-674
Sequence 674, Application US/09973278
Publication No. US20040044191A1
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P2010P2
CURRENT APPLICATION NUMBER: US/09/973,278
CURRENT FILING DATE: 2001-10-10
Prior Application Number: 60/239,899
Prior Filing Date: 2000-10-13
Prior Application Number: 09/227,357
Prior Filing Date: 1999-01-08
Prior Application Number: PCT/US98/13684
Prior Filing Date: 1998-07-07
Prior Application Number: 60/051,926
Prior Filing Date: 1997-07-08
Prior Application Number: 60/052,793
Prior Filing Date: 1997-07-08
Prior Application Number: 60/051,925
Prior Filing Date: 1997-07-08
Prior Application Number: 60/051,929
Prior Filing Date: 1997-07-08

PRIOR APPLICATION NUMBER: 60/052,803
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/052,732
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,931
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,932
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,916
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,930
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,918
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,920
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/052,733
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/052,795
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,919
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/051,928
PRIOR FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: 60/055,722
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,723
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,948
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,949
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,953
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,950
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,947
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,964
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/056,360
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,684
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,984
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/055,954
PRIOR FILING DATE: 1997-08-18
PRIOR APPLICATION NUMBER: 60/058,785
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: 60/058,664
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: 60/058,660
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: 60/058,661
PRIOR FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 947
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 674
LENGTH: 878
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (741)..(741)
OTHER INFORMATION: n equals a,t,g, or c
US-09-973-278-674

Query Match 8.7%; Score 224.8; DB 13; Length 878;
Best Local Similarity 99.1%; Pred. No. 5e-57;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGCATAGTTAAGCCAGCCCGGCAACCCGCTGAGCGGCC 2410

```
Db 202 TGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 261
Qy 2411 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470
Db 262 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 321
Qy 2471 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 2530
Db 322 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 381
Qy 2531 ATACGGCTATTTTATAGTTATGTCATGATATATATAGTTTCTTAG 2578
Db 382 ATACGGCTATTTTATAGTTATGTCATGATATATATAGTTTCTTAG 429

RESULT 10
US-10-092-154-1872
; Sequence 1872, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009CI
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1872
; LENGTH: 878
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741)
; OTHER INFORMATION: n equals a,t,c,g, or c
US-10-092-154-1872

Query Match 8.7%; Score 224.8; DB 15; Length 878;
Best Local Similarity 99.1%; Pred. No. 5e-57; 2; Indels 0; Gaps 0;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2351 TACGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 2410
Db 202 TGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 261
Qy 2411 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470
Db 262 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 321
Qy 2471 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 2530
Db 322 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 381
Qy 2531 ATACGGCTATTTTATAGTTATGTCATGATATATATAGTTTCTTAG 2578
Db 382 ATACGGCTATTTTATAGTTATGTCATGATATATATAGTTTCTTAG 429

RESULT 11
US-09-758-987-6/c
; Sequence 6, Application US/09758987
; Patent No. US20010029014A1
; GENERAL INFORMATION:
; APPLICANT: De Beuckeleer, Marc
; TITLE OF INVENTION: Methods and kits for identifying elite event GAT-ZM1 in biology
; TITLE OF INVENTION: samples
; FILE REFERENCE: 514412-2025.1
; CURRENT APPLICATION NUMBER: US/09/758,987
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 09/481,049
; PRIOR FILING DATE: 2000-01-11
```

```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1073)
; OTHER INFORMATION: Sequence comprising a 5' flanking region of GAT-ZM1
US-09-758-987-6
```

```
Query Match 8.7%; Score 224.8; DB 9; Length 1073;
Best Local Similarity 99.1%; Pred. No. 5.8e-57; 2; Indels 0; Gaps 0;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2351 TACGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 2410
Db 709 TGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 650
Qy 2411 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470
Db 649 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 590
Qy 2471 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 2530
Db 589 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 530
Qy 2531 ATACGGCTATTTTATAGTTATGTCATGATATATATAGTTTCTTAG 2578
Db 529 ATACGGCTATTTTATAGTTATGTCATGATATATATAGTTTCTTAG 482
```

```
RESULT 12
US-09-891-865A-11
; Sequence 11, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHAMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC265
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: M198A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2297
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: Description of Artificial Sequence: cloning vector
; OTHER INFORMATION: derived from pUC18
US-09-891-865A-11
```

```
Query Match 8.7%; Score 224.8; DB 10; Length 2297;
Best Local Similarity 99.1%; Pred. No. 1e-56; 2; Indels 0; Gaps 0;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2351 TACGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 2410
Db 133 TGCTGATGCGCATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACCGGCC 192
Qy 2411 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470
Db 193 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 252
Qy 2471 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 2530
Db 253 TGCATGTGTCAAGAGTTTTCACCGTCATACCGAAACGGCGGACAGAAAGGGCTCGTG 312
```

QY 2531 ATACGCCATTTTATAGTTAATGTCATGATAATAATGTTCTTAG 2578
DB 313 ATACGCCATTTTATAGTTAATGTCATGATAATAATGTTCTTAG 360

RESULT 13

```

US-10-365-062-5/c
? Sequence 5, Application US/10365062
? Publication No. US20030145343A1
? GENERAL INFORMATION:
? APPLICANT: Pfizer Inc.
? APPLICANT: McNeil, John D.
? APPLICANT: Ahlhaugen, Michael K.
? TITLE OF INVENTION: TRANSGENIC ANIMALS EX
? FILE REFERENCE: PC10142B
? CURRENT APPLICATION NUMBER: US/10/365, 062
? CURRENT FILING DATE: 2003-02-11
? PRIOR APPLICATION NUMBER: US 09/496,445
? PRIOR FILING DATE: 2000-02-02
? PRIOR APPLICATION NUMBER: US 60/118,478
? PRIOR FILING DATE: 1999-02-03
? NUMBER OF SEQ ID NOS: 7
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 5
? LENGTH: 2462
? TYPE: DNA
? ORGANISM: Escherichia coli
US-10-365-062-5

```

```
Query Match      8.7%; Score 224.8; DB 15; Length 2462
Best Local Similarity 99.1%; Pred. No. 1.1e-56;
Matches 226; Conservative 0; Mismatches 2; Indels 0;
```

| | | | | | |
|----|------|--|------|--|------|
| Qy | 2411 | TGACGGGCTTGTCTCTCCGGGCATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGAC | 2470 | TGACGGGCTTGTCTCTCCGGGCATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGAC | 2470 |
| Db | 2298 | TGACGGGCTTGTCTCTCCGGGCATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGAC | 2233 | TGACGGGCTTGTCTCTCCGGGCATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGAC | 2233 |
| Qy | 2471 | TGCATGTGTCAAGGTTTTCACCGTCATACCCGAAACGGCGCAGACGAAAGGGCCTCGTG | 2533 | TGCATGTGTCAAGGTTTTCACCGTCATACCCGAAACGGCGCAGACGAAAGGGCCTCGTG | 2533 |
| Db | 2238 | TGCATGTGTCAAGGTTTTCACCGTCATACCCGAAACGGCGCAGACGAAAGGGCCTCGTG | 2172 | TGCATGTGTCAAGGTTTTCACCGTCATACCCGAAACGGCGCAGACGAAAGGGCCTCGTG | 2172 |
| Qy | 2531 | ATAGCCCATTTTATAGGTTATGCAATGATATATATGCTTCTTAG | 2578 | ATAGCCCATTTTATAGGTTATGCAATGATATATATGCTTCTTAG | 2578 |
| Db | 2178 | ATAGCCCATTTTATAGGTTATGCAATGATATATATGCTTCTTAG | 2131 | ATAGCCCATTTTATAGGTTATGCAATGATATATATGCTTCTTAG | 2131 |

RESULT 14

```

US-10-244-142A-2/c
Sequence 2, Application US/10244142A
Publication No. US20030199516A1
GENERAL INFORMATION:
APPLICANT: Moser, Heinz B.
APPLICANT: Baird, Ridon E.
APPLICANT: Buerli, Roland W.
APPLICANT: Ge, Yigong
APPLICANT: White, Sarah
APPLICANT: GeneSocft, Inc.
TITLE OF INVENTION: Methods of Treating Infection by Drug Resistant
TITLE OF INVENTION: Bacteria
FILE REFERENCE: 020891-00910US
CURRENT APPLICATION NUMBER: US/10/244,142A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/322,704
PRIOR FILING DATE: 2001-09-13
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2762
TYPE: DNA

```

```

; ORGANISM: Artificial Sequence
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:double stranded
; OTHER INFORMATION: circular Plasmid B
US-10-244-142A-2

```

| | | | | |
|-----------------------|--------------|--------------------|---------------|--------------|
| Query Match | 8.7%; | Score 224.8; | DB 15; | Length 2762; |
| Best Local Similarity | 99.1%; | Pred. No. 1.2e-56; | | |
| Matches 226; | Conservative | 0; | Mismatches 2; | Indels 0; |
| | | | | Gaps 0; |

| | | | |
|----|------|---|------|
| QY | 2351 | TACCTCATGCGCATGTTAAAGCAGACCCCGACACCCCGCAACACCCGCTGACGCGCC | 2410 |
| Db | 2410 | TGCTCTATGCGCATGTTAAAGCAGACCCCGACACCCCGCAACACCCGCTGACGCGCC | 2351 |
| QY | 2411 | TGACGGGCTTGTCTGCTCCGGGCATCCGCTTACAGACAAAGTGTGACCGTCTCCGGAGC | 2470 |
| Db | 2350 | TGACGGGCTTGTCTGCTCCGGGCATCCGCTTACAGACAAAGTGTGACCGTCTCCGGAGC | 2291 |
| QY | 2471 | TGCATGTGTCAAGAGTTTTCAACCTCATCACCCGAACGGCGGANAAGAAAGGCGTGTG | 2530 |
| Db | 2290 | TGCATGTGTCAAGAGTTTTCAACCTCATCACCCGAACGGCGGANAAGAAAGGCGTGTG | 2231 |
| QY | 2531 | ATACGGCTATTTTATAGTTAATGTCATGATATATTAATGTTCTTGA | 2578 |
| Db | 2230 | ATACGGCTATTTTATAGTTAATGTCATGATATATTAATGTTCTTGA | 2183 |

RESULT 15

```

US-10-244-142A-1/C
/ Sequence 1, Application US/10244142A
/ Publication No. US2003019516A1
/ GENERAL INFORMATION:
/ APPLICANT: Moser, Heinz E.
/ APPLICANT: Baird, Eldon B.
/ APPLICANT: Buerli, Roland W.
/ APPLICANT: Ge, Yigong
/ APPLICANT: White, Sarah
/ APPLICANT: Genesoft, Inc.
/ TITLE OF INVENTION: Methods of Treating Infection by Drug Resistant
/ TITLE OF INVENTION: Bacteria
/ FILE REFERENCE: 020891-009100S
/ CURRENT APPLICATION NUMBER: US/10/244,142A
/ PRIOR FILING DATE: 2002-09-12
/ PRIOR APPLICATION NUMBER: US 60/322,704
/ PRIOR FILING DATE: 2001-09-13
/ NUMBER OF SEQ ID NOS: 20
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 2803
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:double stranded
/ OTHER INFORMATION: circular Plasmid A
US-10-244-142A-1

```

| | | | | |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match | 8.7%; | Score 224.8; | DB 15; | Length 2803; |
| Best Local Similarity | 99.1%; | Pred. No. 1.2e-56; | | |
| Matches 226; | Conservative 0; | Mismatches 2; | Indels 0; | Gaps 0; |

QY 2351 TACGCTATGCCGCATAGTTAAAGCCAGGCCCGCAACCCGCTGAGCGGCC 2410

Db 2420 TGCTCTGATGCCGCAATAGTTAAGCCAGGCCCGCAACCCGCTGAGCGGCC 2361

QY 2411 TGAAGGGGCTTGCTGCTCCGGGCATCCGCTTACAGACAAAGCTGACCGTCTCCGGGAGC 2470

Db 2360 TGACGGGCTTGCTGCTCCGGGCATCCGCTTACAGACAAAGCTGACCGTCTCCGGGAGC 2300

QY 2471 TGCAATGCTCAGAGGTTTTCAACGTCATCAACCGACGGAGACGAAAGGCGCTCGTG 2530

Db 2300 TGCAATGCTCAGAGGTTTTCAACGTCATCAACCGACGGAGACGAAAGGCGCTCGTG 2241

QY 2551 ATAGCGCTAATTTTAAAGGTTAATGTCATGATGAAATATGATTTCTCTAG 2578

Db 2240 ATACGCTATTATTATAGGTAAATGTCATGATPATAATAGTTCTTAG 2193

Search completed: April 30, 2004, 23:19:29
Job time : 1149.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 08:18:22 ; Search time 3218.8 Seconds

(without alignments)
10745.552 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798
Sequence: 1 atgtttgcacaagctctt.....ttcaacctgtttacagagt 798

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_dat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_other:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| C 1 | 434.2 | 54.4 | 10977 | 1 | AE004352 |
| C 2 | 340.6 | 42.7 | 241900 | 1 | AP005337 |
| C 3 | 339 | 42.5 | 306147 | 1 | AE016805 |
| C 4 | 282.8 | 35.4 | 303450 | 1 | AP005085 |
| C 5 | 42 | 5.3 | 1141 | 6 | AX083744 |
| C 6 | 40.2 | 5.0 | 146953 | 2 | AC138591 |
| C 7 | 40 | 5.0 | 2067 | 3 | AF146758 |
| C 8 | 39.2 | 4.9 | 16237 | 2 | AC012853 |
| C 9 | 39.2 | 4.9 | 74085 | 3 | AC012650 |
| C 10 | 39.2 | 4.9 | 108350 | 3 | AC011761 |
| C 11 | 39.2 | 4.9 | 302540 | 3 | AE003571 |
| C 12 | 39 | 4.9 | 128563 | 2 | AC135793 |
| C 13 | 39 | 4.9 | 168422 | 2 | AC139922 |
| C 14 | 39 | 4.9 | 187063 | 2 | AC107808 |
| C 15 | 38.8 | 4.9 | 2669 | 2 | AX125495 |
| C 16 | 38.8 | 4.9 | 304680 | 1 | AE017002 |
| C 17 | 38.4 | 4.8 | 303250 | 1 | AP003193 |
| C 18 | 38.2 | 4.8 | 1318 | 9 | BC022446 |
| C 19 | 38.2 | 4.8 | 5533 | 9 | HSN807894 |
| C 20 | 38.2 | 4.8 | 5841 | 9 | HSN803639 |
| C 21 | 38.2 | 4.8 | 6242 | 9 | HSN807593 |
| C 22 | 38.2 | 4.8 | 110146 | 2 | AC093684 |
| C 23 | 38.2 | 4.8 | 160142 | 9 | AX655393 |
| C 24 | 38 | 4.7 | 2000 | 6 | AX655393 |
| C 25 | 37.6 | 4.7 | 2000 | 6 | AX655393 |
| C 26 | 37.4 | 4.7 | 1141 | 6 | AX083744 |
| C 27 | 37.2 | 4.7 | 189271 | 2 | AC129309 |
| C 28 | 37 | 4.6 | 17731 | 2 | AC129133 |
| C 29 | 37 | 4.6 | 187121 | 2 | AC112418 |
| C 30 | 36.8 | 4.6 | 115762 | 2 | AP000663 |
| C 31 | 36.8 | 4.6 | 155332 | 2 | AC109581 |
| C 32 | 36.8 | 4.6 | 172830 | 2 | AP001320 |
| C 33 | 36.6 | 4.6 | 186955 | 9 | AP000848 |
| C 34 | 36.6 | 4.6 | 4894 | 9 | HSN807890 |
| C 35 | 36.6 | 4.6 | 221640 | 2 | BX322637 |
| C 36 | 36.4 | 4.6 | 82073 | 9 | HS172K10 |
| C 37 | 36.4 | 4.6 | 168253 | 2 | AC068527 |
| C 38 | 36 | 4.5 | 73465 | 9 | HS256M13 |
| C 39 | 36 | 4.5 | 161514 | 9 | AC078845 |
| C 40 | 36 | 4.5 | 166239 | 9 | HS214G14 |
| C 41 | 36 | 4.5 | 253339 | 2 | AC112075 |
| C 42 | 36 | 4.5 | 273908 | 2 | AL162851 |
| C 43 | 36 | 4.5 | 340000 | 9 | HS21C004 |
| C 44 | 35.8 | 4.5 | 254050 | 3 | PFA929358 |
| C 45 | 35.6 | 4.5 | 552 | 6 | AR345675 |

ALIGNMENTS

| RESULT 1 | AE004352/c | 10977 bp | DNA | 1 linear | BCT 10-OCT-2003 |
|------------|---|----------|-----|----------|-----------------|
| LOCUS | AE004352/c | | | | |
| DEFINITION | Vibrio cholerae O1 biovar eltor str. N16961 | | | | |
| ACCESSION | AE004352 | | | | |
| VERSION | AE004352.1 | | | | |
| KEYWORDS | GI:9657475 | | | | |
| SOURCE | Vibrio cholerae O1 biovar eltor str. N16961 | | | | |
| ORGANISM | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio. | | | | |
| REFERENCE | 1 (bases 1 to 10977) | | | | |
| AUTHORS | Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, R.K., Peterson, J.D., Umayam, L., | | | | |

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Gill,S.R., Nelson,K.E., Read,T.D., Tetteijn,H., Richardson,D.,
Emoiaeva,M.D., Vamathavan,J., Base,S., Qin,H., Dragoi,I.,
Sellers,P., McDonald,L., Uterback,T., Fleishmann,R.D.,
Niernan,W.C. and White,O.
DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
cholerae
Nature 406 (6795), 477-483 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 10977)
Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
Gill,S.R., Nelson,K.E., Read,T.D., Tetteijn,H., Richardson,D.,
Emoiaeva,M.D., Vamathavan,J., Base,S., Qin,H., Dragoi,I.,
Sellers,P., McDonald,L., Uterback,T., Fleishmann,R.D.,
Niernan,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
Nekalans,J.J., Venter,J.C. and Fraser,C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..10977
/organism="Vibrio cholerae O1 biovar eltor str. N16961"
/mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:243277"
/chromosome="III"
/note="biotype: El Tor"
complement (81..974)
/gene="VCA0090"
complement (81..974)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF96004.1"
/db_xref="GI:9657476"
/translation="MSLVNLTQLCLAIWGSFLEMRILAAHSFGAVILIRAVFPA
LSFLVAILRRSLPRHWPPIILGINTVPLLPAYVALTLNVAITLSTAP
INCAVIGPLMHGPPSRKAVAGLLGVSVATVGMVAIGHAALMVCALAAAS
YGLATVTKQAPQLSAFENAHSGMWACLVAPLWVFPLRTPSLMGAITLGLV
CTGLAVILYRLVKAIGASTSVTLIPVGLIWMGYILDEPILGTLTGLTLVAG
TMLVTGFSLRNARLSROPOPH"
1204..1500
/gene="VCA0091"
1204..1500
/gene="VCA0091"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF96005.1"
/db_xref="GI:9657477"
/translation="MARPKIPRIECPSPASCFKXNGVPIRQLARVELAPDELELR
VQLGLQOQOQALQWVSROTLAVKARFVVDCLHQALYIQALDKSSD"
complement (1671..2300)
/gene="VCA0092"
complement (1671..2300)
/gene="VCA0092"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96006.1"
/db_xref="GI:9657478"
/translation="MKIVTIGLWMEGLVMOPLAQLNKLAGYFVLSYNTLAIIDEX
VQOTLSALADPRINVTGSHIGMIGWGLVSPNVISHVVALASPLKASIVP
KIQQLGMLGNALHYGLHQDSWELPORGCTAGTLRGRFRILLGSGMKCDGY
TVAETISGTHLLHOSHGLVTSKTRQIDYFIRNDFQHKIPE"
2400..2573
/gene="VCA0093"
2400..2573
/gene="VCA0093"

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2400..2573
/gene="VCA0093"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96007.1"
/db_xref="GI:9657479"
/translation="MEQPCWLRVNPITWFKYVNEDEPCTLLPYTPSRLLAKGLEKXH
DDCLFKRSQDLN"
2801..3157
/gene="VCA0094"
2801..3157
/gene="VCA0094"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF96008.1"
/db_xref="GI:9657480"
/translation="MDVNIKSPRIPIINDVVALQQLERISKWTHTLIMHDSSESARF
GEIQSLANGISAKMLSRMLKMBEIGWVRQQTIPVYVYGLTSGRVVANGLYRW
IDHVENEREPLLESN"
3426..4124
/gene="VCA0095"
3426..4124
/gene="VCA0095"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96009.1"
/db_xref="GI:9657481"
/translation="MYOMKLSRWLPLSPVLPFLIGLSLFTVSTSPRTNLPDEROE
VIAKPAPELRSIHHPAVELSTENRVNVEGEDEGEAPRVLEPONTLRQWYPI
NTPBELADLPKQAVTDIOLATFENITDQRYVMAEDSATLPIPNESIRVILDEQ
QAQSIQTMAGNFEFGTFPVYTFGQSHMGITGHPQSIKIEVGHQAWYQVDS
HGFEHDTQSDRHH"
4167..5390
/gene="VCA0096"
4167..5390
/gene="VCA0096"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96010.1"
/db_xref="GI:9657482"
/translation="MKKASALAVSGLCSPAIMATTEIDVGLYTPDPAKGRKQEHV
AQWQHSYVANKRYLKDSGLDKVNLAKTSKVOITDTPGLKSSQSEVLDPATPNRIDP
AFADVEAYRQVAGDVAALFRYLDVNNSPDYERQPNQYSISGLMVAIPSAWKTPQ
NAKXSYSHSYLNEGCAETFIHELGNFGNAHAEQYRELPHNNQTEADVAGYKIG
QFATIMAYPHLFGVGRSYKSSPNLQCEGAPCGVDYASVRAIGLTAHIAQVYGT
KPPVDQNPQDTEPTDNTNNVFTIKGPLALPDKKILTLPIVSAQSSAQVADITH
EYRGDSIRLFAFDPGDSYVWLKQANRYDRGGSYVGTLDVDPSSAEGEWRLBIQHF
GKLGTLNQOQITPP"
complement (5403..6119)
/gene="VCA0097"
complement (5403..6119)
/gene="VCA0097"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF96011.1"
/db_xref="GI:9657483"
/translation="MTVITDMICLRAPKGIETLLVYKSNPNRPGCLMAIPGWVFD
EDLSAQGEAPADPDQSRARRICRKHITVYENISDPLVDNGKRPSPSSISISYA
LHNSVYTOLEBEMDKSRVNRVAFETIIOGCOVLAFLVAGIOHAKMLRAAVETYS
VVLPSLEETLVADITLBAKRGVDVNRVTRARLDITGVIVYSTNNLAASSKSGKGP
ATVYRLASNEVITFQCLNG"
6518..7825
/gene="VCA0098"
6518..7825
/gene="VCA0098"

CDS
5518..7825
/gene="VCA0098"
/note="similar to GB:J05568 SP:P18133 PID:147307 GB:U00096
PID:1651456; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="nucleotide phosphoribosyltransferase"
/protein_id="AAF96012.1"
/db_xref="GI:9657484"
//translation="MNPRLFSFPHIRSLDDDAKXINMOAIIHFPDVSRYLLIR
SEDSGLDADROIHAHGLRFSADIDHITQAPHLKATQSLKRFHFVQEV
EMGIVKQGGKQRLIRISGSRWDTLLVETLVAIVSEVRSGQNAEVPADIPVLT
KIDQAKIERGINNPSLITWETERRSSQVORDVLAICQEIPOWLTGSNHFAR
BEDLPIGTIAHEWFMGQALYNERDSQVALERWLAFDGMALAPDITLIDAFIA
DENRLANADGVRHDSGCPFMGDKMIAHYQGLDPTTCLFIFSGLDPDALERC
EYRGRKVISRGISGELTNLANWNPANAGVRRPISYIKIAECGRVAKISDQPERK
AMCEDPFLANLKRFPNIELVDALIQELRHOKSPRAYISA"
7856..8896
/gene="VCA0099"
/note="similar to GB:U00096 PID:1742678 PID:1742686
PID:1742694 PID:1742718; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, G6o/Idh/MocA family"
/protein_id="AAF96013.1"
/db_xref="GI:9657485"
//translation="MNPRLPKKGVIGYKSAKTHLPIDTLESITLSISSQQQV
QEDWFOIAFDSQRLITSDVDVITTAIPNVHPLKALLEGKHYVERPVTQI
EQGRILBLAQGGLLSEFNRNRWDGDFLVKXILEGQGLGVKVFESHPDRPVV
RORHBOAQEGGILFDLAPHLIDQALFGLPQSLSDSCMDPDAITLIDPLQLY
YQHVRLAHANLYSPENVRVYGLSLKRYGLDPEDELKXGERTHQWQEMP
EYGLIHAEGENIVTELGYQHFTQVVEAIRGAKNPVSASALQSLILALE
SSAKGRLAATL"
complement(8933..9304)
/gene="VCA0100"
complement(8933..9304)
/gene="VCA0100"
/note="similar to GP:7226016; identified by sequence
similarity"
Query Match 54.4%; Score 434.2; DB 1; Length 10977;
Best Local Similarity 72.5%; Pred. No. 8.2e-117;
Matches 576; Conservative 0; Mismatches 218; Indels 1; Gaps 1;

CDS
5841 TTGAGCAATCAGATTTCTACTATGCCCTGCTCAATCATTCACAGCAATGGA 5782
3 GTTGTGCACAAAGCTTTATTTACATTTGACATCATGCACTTACCTAAATAAGCC 62
6200 GTTAATCACATTTGTATTTACATTTGACCTTACCATGATGATAGGAT-AGCG 6142
63 GTTGTATTAAGGAAACCATATGATGATCTATGATGATGATGATGATGATGATG 122
6141 CACCTTAATGATGAAACCTTGTATGATGATGATGATGATGATGATGATGATG 6082
123 GCGGAATATATCCAGGTTTACTGCTGAAGCGCTTATCCAAATCGCGCAGATGATG 182
6081 CCCCAAGGCAATGAACTTGTGCTCAACGCGCAATCCAAACCGCTGATGATGCG 6022
183 TAAATGGGCAATGCTGCGCGGATGATGATGATGATGATGATGATGATGATG 242
6021 TCTTTGGGGATTCCTCGGATGATGATGATGATGATGATGATGATGATGATG 5962
243 ACTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 302
5961 ACCCGGTGCAAGATTTGATTCGCGCGGATGATGATGATGATGATGATGATG 5902
303 TTAATCTAATTTATCAGGATCGGCTGATGATGATGATGATGATGATGATGATG 362
5901 TTAACCACTTCATAGATGATGATGATGATGATGATGATGATGATGATGATG 5842
363 TTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 422
5841 TTGAGCAATCAGATTTCTACTATGCCCTGCTCAATCATTCACAGCAATGGA 5782

423 AGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 482
5781 AGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5722
483 AGAATGCCCGCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 542
5721 TAAACAGATTCGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 5662
543 CGCTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 602
5661 CGAGCGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5602
603 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 662
5601 GCACATTTAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATG 5542
663 CGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 722
5541 TCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5482
723 CAAGAGCGCAACAGCAACGCTTATGATGATGATGATGATGATGATGATGATG 782
5481 TPAAGGCGGCAACCTGCGACGCTTATGATGATGATGATGATGATGATGATG 5422
783 AACCTGTTACGAGG 797
5421 GACTGCTTCGCTG 5407

RESULT 2
AP005337/c 241900 bp DNA linear BCT 04-DEC-2003
Vibrio vulnificus YJ016 DNA, chromosome I, complete genome, section
8/14.
AP005337 BA000037
AP005337.1 GI:37198635
Vibrio vulnificus YJ016
Vibrio vulnificus YJ016
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
AUTHORS
Chen,C.Y., Wu,K.M., Chang,Y.C., Chang,C.H., Tsai,H.C., Liao,T.L.,
Liu,Y.M., Chen,H.J., Shen,A.B., Li,J.C., Su,T.L., Shao,C.P.,
Lee,C.T., Hor,L.I., and Tsai,S.F.
TITLE
Comparative Genome Analysis of *Vibrio vulnificus*, a Marine Pathogen
Genome Res. 13, 2577-2587 (2003)
2 (bases 1 to 241900)
AUTHORS
Chen,C.Y., Wu,K.M. and Tsai,S.F.
DIRECT SUBMISSION
Submitted (29-MAY-2002) Shih Feng Tsai, National Health Research
Institutes, Division of Molecular and Genomic Medicine; 128,
Yen-Chiu-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China
(E-mail:petesai@nhi.org.tw, Tel:886-2-8146-1041,
Fax:886-2-2789-0484)
COMMENT
This sequence was determined by the Sequencing Core of the National
Yang-Ming University Genome Research Center (YMC;
http://genome.ym.edu.tw).
FEATURES
source
1..241900
/organism="Vibrio vulnificus YJ016"
/mol_type="genomic DNA"
/strain="YJ016"
/db_xref="taxon:196600"
/chromosome="I"
197..1024
/gene="VVJ706"
197..1024
/gene="VVJ706"
/codon_start=1
/transl_table=11
/product="formate dehydrogenase accessory protein"

```
/protein_id="BAC94470.1"
/db_xref="GI:37198636"
/translation="MKPKVITKSNPLQTEVEVPEDEYGEKLKIQACSRPLTVLNL
KWEVTLMTLGSREPALVIGYIKKNSFLSEVSLSVIIDMETNSAAVYTKENVHIE
QALKKTVISGCGQVTMYGNVKNQULHAYVPQVTLKQSEITKTLTHLTYNDYTKVAG
AVHGCACVCKGDEVRSEFVEDVGRNNAVDTLAGEMLNDESGADKIFTYTGRLTSVAVIK
VWQMGPVLLSSSGVQMGMDLARGQITTTIARAGLRFQVPTGGENTIIFFVKQGES"
1151..2908
/gene="VVI707"
1151..2908
/gene="VVI707"
1151..2908
/codon_start=1
/translation_table=1
/product="methyl-accepting chemotaxis protein"
/protein_id="BAC94471.1"
/db_xref="GI:37198637"
/translation="MSGSPVQKRFFPSRSIRAKLLINVTLLGVAFAVYENNSAKH
LSSLERASTENSSSIDLMLRHEKDYIARSEKRLTSEFATYSQLOQRIDHLSDL
ASSSIAIDQROQIVTTLQOYQFOQLAQLANDALQSEDLFAARSMUKTNVLSAND
GELTAQFLDLHDLTFTVEPTQEHKVALLEGLSQAKSEADLPATFDYSVTLRSYV
SAIETGLTANEGELRGALRSNVKTEBOATITDQTTIKQGEQTADIRSHQOMGVVI
VLLSSLVYIGRSILSRIRKAINILMDIDANGSGDLYVMNAGSDELAOLSRDLF
TSHLHNITKOASVMDVLCSSSSSROAORSMONARVNSRADKALQLTATGCRID
TLASTIEBOAQIQALPAQSRRENNVSAIQIAECTNLTALNATIBARAGSEGRG
AAVADVRQLSRLTNDSTQISTVQALVKGIAQVKNKADSDPARFTNDTRVVIS
ALRGISTQITEMPDNTQIATASBESQSWASBIDRNTQIADLADTYEIVSGSVCS
EQVDVSHLEKTVAAQFKY"
complement(3018..4610)
/gene="VVI708"
complement(3018..4610)
/translation="VVI708"
/translation="containing duplicated ATPase domains"
/codon_start=1
/translation_table=1
/product="ABC-type transport system, ATPase component"
/protein_id="BAC94472.1"
/db_xref="GI:37198638"
/translation="MISTANTIQGAKPLPENISVKGSGNRVGLIGANGCGKSTFM
KILSGLEPTSGVSYDPNERVAKLNQDQFAVEFTVIDVTIMGHELMKVCQEDRI
VSLPENSEBQKRVADLEVEFPMDMGYMAKAGALLAVGI.PMEQVGLMSVAPG
KRLVLAQILFADPDIIMLDEPTNNLMDTIRLRETTLNARCTMIISHDHPLANSV
CTKMDLDDGELVPGNDEVTWATASQARELLSPNACKQIAELCTQVAFPSANA
SKKAKATSAKQIDKIKLDEVASRRONPFRPESKELFRNALIVENISQEFEDFL
SDPNAIPVEGRVAILTGENGVCTTILNTLACVLEPRTGEPKMSNANIGYACQDAH
EPFADNLMEWQMRQEGDDQVVRGFLGRVLFQDDITKSKVLSGEGQRMILGK
LMKHPNMLMDSPINAMDESIENNALEQKTLFFVSHDRVFDLSLAVRIIEIR
DKGITPFKGYSEYLSKRGVES"
4834..7179
/gene="VVI709"
4834..7179
/gene="VVI709"
/translation="VVI709"
/translation="identified by GeneMark and Glimmer2"
/codon_start=1
/translation_table=1
/product="hypothetical protein"
/protein_id="BAC94473.1"
/db_xref="GI:37198639"
/translation="MOKLRLFYCPTLVAPATPROMQYAAQTPAQVSAKALSTESDL
SALDPLFMGPKEHIRVNDLSAVLELQTEGDAFALAOYKAPLAPAOVLWLNQY
LTLNLSGSKARLQDLTASRQLTGFGVGTQORBELATKNIETLVAQVLFQSRF
RALIKNTFISPVIVITGLKFAVYLLVMMNANSKRITLTPRESQLENNIPLMVK
FIWYISRANKAIAIMLAIITVSAIILASLPSLOHLYLEIFTWIILGSGSIAISPILEPA
VYNSRSSKQVVALRELTATRRVYWSAIVAGVLIQSVRTIGGTIYYWVLIISHPK
LVYVYLRKMKRKYFEVVDRIETRPVWOMAIKQDSFTLIPATAIYAVMLISHPK
HRIATISNTTMSQALAIPIREYAKQSNQDNDLSVIRIGDAFOVYVPEQDSE
WIDYASDEIQNLSRYLLTDSFALCVLSRSGVTSVLKAILKVRNRIPLTANPYD
GYBELKQALATGLAEESTEQVILAFLKRSRTPYLAIVDNRQLVKPEVGLAQLMK
LTLNLRSKQNHVVVIAIEKASVRFVDRARGRLFDWVSLPRNNEXQISALDSRFA
NODIENLSPDGLVVPKQDDETSSEERAKQGFYLIWHYGDGNTVALISFRFLSLR
RKNSDSVVVRLFAPESEKLEQMPKMLAVRSIYOLEVAKPEELSECTQSIABVI
GTRPFBSRGFTIMNEDKARVSDHMRIRYITVLDNRHLLVK"
7181..8062
/gene="VVI710"
7181..8062
/gene="VVI710"
/translation="VVI710"
/translation="small-conductance mechanosensitive channel"
/product="VVI710"
/protein_id="BAC94474.1"
/db_xref="GI:37198640"
/translation="NNKKYALFAALISAPAFTEIANMENIQOPASLIRNMGVFF
SVIIATITWLLKFNNSVYESIGQFOYKMLQKQSPFQFFIYLAGLVPMGSR
INDHIALIGGLTAVSGFALKDLAASRIGLTVMDIPROVGDRTFEGNGDITAI
GLRSVTRKRTITDITITTPNNKFLNBYTSVSNYALDMQVITPFYVWDEDTYARELI
QEAASSRITHLRPPVVALVKQTTNDNYALIQLTCAVVDVAYEKLFESDITLRVWK
EFKKGKXPRPIRPSQ"
complement(8150..10276)
/gene="VVI711"
complement(8150..10276)
/translation="VVI711"
/translation="VVI711"
/codon_start=1
/translation_table=1
/product="signal transduction histidine kinase"
/protein_id="BAC94475.1"
/db_xref="GI:37198641"
/translation="MAINRRRYVLIAMHAKMSANTKSGEKRRIMQSDOTLMSWRY
RKSMAVYRLITFSAPITFTLALIGISVYNSIHTWSALIDVSRGLVANSSTLR
LQKQANVYAFABSTQFRTRIQPMQDDQLQMVHKKQYQULDEAFRVDITBEK
FRFLDLTKRSFPDVLSFELQRLDPLAPRAVPLIKTKETRGILSTVLFENQ
SNDLIGLIDGGLLNNSTVLDQIRDLVIGDEDSRPIGTVLFMEDLRSVSTVPLD
SEDOIIRALITRVSGEVHKSVLQQGERVWRAVVDAMVITAYQIRDOYNNITGLY
TGYLMPFVKAVALTNIYEISVITLTLVSGGLVYSGRDLFPRIRKHVYLVQVQ
KDKRIGLIGDENHELAQAKOPDNMLDKTKEEQOKANLEBOKVETASLBEK
TEQLEHRIIRLNRDCLTNEKALAGETAGIARHINNPVAVIIGVNLIFPELGE
DAAREEIEFTVAQIDRLNITRSLQISQSGQVDEITWQVNPITESTITLVKTS
SKKRTIEFTVELNHNSTSEVENRQQLQIVNLQMANLHANNGBRLTISEDLWEGE
SVGAVIHIEDQGGIGIKPEKLSRIFFPEYTYKBDGTGLSVSQSLLSQSGELIKVRSQ
LGISTSVYLPKRNKAKLDTLLIQ"
complement(10173..11678)
/gene="VVI712"
complement(10173..11678)
/translation="VVI712"
/translation="VVI712"
/codon_start=1
/translation_table=1
/product="sigma-54 dependent response regulator"
/protein_id="BAC94476.1"
/db_xref="GI:37198642"
/translation="MIMHTSYSTSELESMSIQDLPNSVQYAPSVLYVNDYGVN
QALIKKALGKSPFSVADVSIIEQAEKRVNHYDLVLDNLNFGSGITEBERFNPBE
RKDVIVIMTYADLETAITALKIGATDQFLKPNLQMQAVRCKDRVLNARMQYAL
KRDQVNIHEMVGNSDKTRQLKMITQFIPSKACVYLIEBESGKELVARGVHLASG
RTGPFVPLNCGALAPBELSEELFGHTSGAFTGAKKAREGLFVALIGTFLDEIGEMP
LAWOSSILRVLEORSIRPVSGEKEIPVDVAVVATNRDLOQEVKGFREDLYRLNV
LKTIEVPLRERKADLOLVPFTKQSGELAMAPQAHEDITAMGDYSGVGRLEIR
NLIERCILLGPPPAHTREINSGMRLNPNVSIYVSHSABVPTPEGKEKFTGYGVPRNDW
TLKVERAHIQQVHVHYDGNKSAASRDLGVARKTLEKRYKEMETBEADVAE"
11942..12760
/gene="VVI713"
11942..12760
/gene="VVI713"
/translation="VVI713"
/translation="VVI713"
/codon_start=1
/translation_table=1
/product="ABC-type tungstate transport system, permease component"
/protein_id="BAC94477.1"
/db_xref="GI:37198643"
/translation="MNNVALTLAATSTTVSFSFSADVHRVALATTTSTYSGULD
YLLEPEKDGKYVDVLAAGTGSIKKNGGVLDVWHTAPKXAEATFVEGYVLPK
LMVNDFTVNGPKADPAVADQNSVLREPKAIADNNVTFTSRGDSGTHKEKELIWAOT
KMEVNFSGYSVQCGKPTLNMASMGITLTTRGTLATNNKLDLAVIPEGOKLEN
PYQVILNPERKRYINIQAKASDWLNVBRGGLINSFRLNGQLFVANADQX"
12767..13465
/gene="VVI714"
12767..13465
/gene="VVI714"
/translation="VVI714"
/translation="VVI714"
/codon_start=1
```

```
CDS
7181..8062
/gene="VVI710"
/codon_start=1
/translation_table=1
/product="VVI710"
/protein_id="BAC94474.1"
/db_xref="GI:37198640"
/translation="NNKKYALFAALISAPAFTEIANMENIQOPASLIRNMGVFF
SVIIATITWLLKFNNSVYESIGQFOYKMLQKQSPFQFFIYLAGLVPMGSR
INDHIALIGGLTAVSGFALKDLAASRIGLTVMDIPROVGDRTFEGNGDITAI
GLRSVTRKRTITDITITTPNNKFLNBYTSVSNYALDMQVITPFYVWDEDTYARELI
QEAASSRITHLRPPVVALVKQTTNDNYALIQLTCAVVDVAYEKLFESDITLRVWK
EFKKGKXPRPIRPSQ"
complement(8150..10276)
/gene="VVI711"
complement(8150..10276)
/translation="VVI711"
/translation="VVI711"
/codon_start=1
/translation_table=1
/product="signal transduction histidine kinase"
/protein_id="BAC94475.1"
/db_xref="GI:37198641"
/translation="MAINRRRYVLIAMHAKMSANTKSGEKRRIMQSDOTLMSWRY
RKSMAVYRLITFSAPITFTLALIGISVYNSIHTWSALIDVSRGLVANSSTLR
LQKQANVYAFABSTQFRTRIQPMQDDQLQMVHKKQYQULDEAFRVDITBEK
FRFLDLTKRSFPDVLSFELQRLDPLAPRAVPLIKTKETRGILSTVLFENQ
SNDLIGLIDGGLLNNSTVLDQIRDLVIGDEDSRPIGTVLFMEDLRSVSTVPLD
SEDOIIRALITRVSGEVHKSVLQQGERVWRAVVDAMVITAYQIRDOYNNITGLY
TGYLMPFVKAVALTNIYEISVITLTLVSGGLVYSGRDLFPRIRKHVYLVQVQ
KDKRIGLIGDENHELAQAKOPDNMLDKTKEEQOKANLEBOKVETASLBEK
TEQLEHRIIRLNRDCLTNEKALAGETAGIARHINNPVAVIIGVNLIFPELGE
DAAREEIEFTVAQIDRLNITRSLQISQSGQVDEITWQVNPITESTITLVKTS
SKKRTIEFTVELNHNSTSEVENRQQLQIVNLQMANLHANNGBRLTISEDLWEGE
SVGAVIHIEDQGGIGIKPEKLSRIFFPEYTYKBDGTGLSVSQSLLSQSGELIKVRSQ
LGISTSVYLPKRNKAKLDTLLIQ"
complement(10173..11678)
/gene="VVI712"
complement(10173..11678)
/translation="VVI712"
/translation="VVI712"
/codon_start=1
/translation_table=1
/product="sigma-54 dependent response regulator"
/protein_id="BAC94476.1"
/db_xref="GI:37198642"
/translation="MIMHTSYSTSELESMSIQDLPNSVQYAPSVLYVNDYGVN
QALIKKALGKSPFSVADVSIIEQAEKRVNHYDLVLDNLNFGSGITEBERFNPBE
RKDVIVIMTYADLETAITALKIGATDQFLKPNLQMQAVRCKDRVLNARMQYAL
KRDQVNIHEMVGNSDKTRQLKMITQFIPSKACVYLIEBESGKELVARGVHLASG
RTGPFVPLNCGALAPBELSEELFGHTSGAFTGAKKAREGLFVALIGTFLDEIGEMP
LAWOSSILRVLEORSIRPVSGEKEIPVDVAVVATNRDLOQEVKGFREDLYRLNV
LKTIEVPLRERKADLOLVPFTKQSGELAMAPQAHEDITAMGDYSGVGRLEIR
NLIERCILLGPPPAHTREINSGMRLNPNVSIYVSHSABVPTPEGKEKFTGYGVPRNDW
TLKVERAHIQQVHVHYDGNKSAASRDLGVARKTLEKRYKEMETBEADVAE"
11942..12760
/gene="VVI713"
11942..12760
/gene="VVI713"
/translation="VVI713"
/translation="VVI713"
/codon_start=1
/translation_table=1
/product="ABC-type tungstate transport system, permease component"
/protein_id="BAC94477.1"
/db_xref="GI:37198643"
/translation="MNNVALTLAATSTTVSFSFSADVHRVALATTTSTYSGULD
YLLEPEKDGKYVDVLAAGTGSIKKNGGVLDVWHTAPKXAEATFVEGYVLPK
LMVNDFTVNGPKADPAVADQNSVLREPKAIADNNVTFTSRGDSGTHKEKELIWAOT
KMEVNFSGYSVQCGKPTLNMASMGITLTTRGTLATNNKLDLAVIPEGOKLEN
PYQVILNPERKRYINIQAKASDWLNVBRGGLINSFRLNGQLFVANADQX"
12767..13465
/gene="VVI714"
12767..13465
/gene="VVI714"
/translation="VVI714"
/translation="VVI714"
/codon_start=1
```


gene
CDS
/locus_tag="VV12372"
/complement (3584..4897)
/locus_tag="VV12372"
/locus_tag="VV12372"
/note="COG1488"
/codon_start=1
/transl_table=11
/product="Nicotinate phosphoribosyl transferase"
/protein_id="AA010746.1"
/db_xref="GI:27361840"
/translation="MGAPLIPAPAIQSVLDLVYKINMGAAYRPOTORYELIV
SDNLSLDVBEVEREINRLAERPDAAQAYIAEKAPITAEFLSYLTERPHPOQV
SVGIFRTAQDCQRLRTINGIWHETILYETLVMSITSELRRKRYAQIPQQLHVL
DEKLDLSBLKRRTITNFRSMBGRRRSPAAQKTMLDVLAARVPELLLTSTNHLA
OENFLPTIGTVAHEMTMAQALVALIQHSQVALDWLEAFNGSLGIALDTIGDAPL
SPDLDKATVAGVRHDSGSPVWGDKIAHESGIDPTEKTLIPTGLDFARALDI
CEYFAGRAQISFGIGFTLDMGNMTNDGTGYPISVYVYKAECSGSPVAKISDEPE
KMCEDIPFLMKQRFGLVDLDAIEFLKMKQCKKRIQSYA"
5176..5889
/locus_tag="VV12373"
/locus_tag="VV12373"
/locus_tag="VV12373"
/note="COG1051"
/codon_start=1
/transl_table=11
/product="ADP-ribose pyrophosphatase"
/protein_id="AA010747.1"
/db_xref="GI:27361841"
/translation="MIVTIDMTCLKEGKLEVLITKTPNRPQHGMSIPGGEVFE
HDSQGGQPADPDSARRICROKITHYPISEPVWGDKPQDPMGASTLAHYA
LHNTNIOIIONGICDQOLAMPALLESVIGEVALLAPHADILRLAKLALEYTS
VHFAEKRFVLSDTIHAYAKGVVNRRTIKRLIDGVIVSANKIKASTTKGKGR
ADVYSLADKHVYFOTCR"
5902..6546
/locus_tag="VV12374"
/locus_tag="VV12374"
/locus_tag="VV12374"
/note="similar to nicotinamide; COG1335"
/codon_start=1
/transl_table=11
/product="Amidase"
/protein_id="AA010748.1"
/db_xref="GI:27361842"
/translation="MYDIPKNTASLDVDPQGFSELCNPELVSGALEIYEAIKONH
TRAKLTLSRDMHPRGAMAEATPRAMLEPVGLPEYDIKMPHCIVGQVILLPLP
ALREYFQMKGIDPDHPYGAETHDQUTSTGGIEFLKARITIVLVGGLADFCV
KXSIQLVADPFVILNLASTRAVPENNPVVSHLTALGVICDDVQLOLHE"
6962..7249
/locus_tag="VV12375"
/locus_tag="VV12375"
/locus_tag="VV12375"
/note="COG0776"
/codon_start=1
/transl_table=11
/product="Integration host factor, alpha subunit"
/protein_id="AA010749.1"
/db_xref="GI:27361843"
/translation="MALTAKDLAENLFELGFSKDKADQTVVFEETIKALENGY
KLSGFGNPRDKNERPGRNPKTGEDIPITARVYVTFRRGQCLKARVENLK"
complement (7347..8123)
/locus_tag="VV12376"
/locus_tag="VV12376"
/locus_tag="VV12376"
/note="COG2357"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AA010750.1"
/db_xref="GI:27361844"
/translation="MSLFLRTTALMLVLVSRAPAPAIPTASNOEKORETSONEVCS

gene
CDS
/locus_tag="VV12377"
/complement (3413..9066)
/locus_tag="VV12377"
/locus_tag="VV12377"
/note="COG0500"
/codon_start=1
/transl_table=11
/product="Thiopurine methyltransferase"
/protein_id="AA010751.1"
/db_xref="GI:27361845"
/translation="MDDAEWKNKASNOIGFLEVDVNPDLRFRGALAPKEETLV
PLCGRTDLATLTKEDHVGASLSLIARSPFAEHFVPTVYPISGHEHLYOFDLS
IYAGDEFTAPLSADIYDRAALIALPEMRVREYVIRIGLNPGRSILLVSDYPOQ
ENAGPPSVYOEIEHFLFAGMYTRLQCOIADHHKPKIAKQIGRSFSEYVYENDK"
complement (9119..10294)
/locus_tag="VV12378"
/locus_tag="VV12378"
/locus_tag="VV12378"
/note="COG0027"
/codon_start=1
/transl_table=11
/product="Phosphoribosylglycinamide formyl transferase 2"
/protein_id="AA010752.1"
/db_xref="GI:27361846"
/translation="MFGATRESATVLLIGSELEKEVAIECORGLVIACTDRYP
APMQVARSYVDFMDASLEKVIABEPFAVPEIEAIDTKLELEBQGLNVPS
AKATKLTNNKNEGIRLQAEELGITSPPADYVQOFNAVEVLSIPCVKPPSSG
KQGVISPADIEKARQAQVQEGRTAGYVGEFIDPYEITLLPRAVDGVHPCAP
IGRHOEDVDRESNOPOQMSENAIIKAEVTAQVNAALGGYGFVLELFGKDGFTN
EVSPPHNDGTLLTISOEMSEFALHVAFTGMVKNITVOYGPSASVITIGNGOENLR
PDGMSDLBQPOQLRFGKPDINGRRRLGVITRRSSTREKAVDAIEBAKIKITV"
complement (10426..11334)
/locus_tag="VV12379"
/locus_tag="VV12379"
/locus_tag="VV12379"
/note="COG0295"
/codon_start=1
/transl_table=11
/product="Cytidine deaminase"
/protein_id="AA010753.1"
/db_xref="GI:27361847"
/translation="MQLQFTMKSRFENALASAPESLSHAPILIIAADFASLSTVO
FDLHKOTGMDNLQVALLPFAAYVAPISSEFYGAVRGLSALYGANMEFQV
OLGQTYARQSAISHAMMGEOGLSDITINPSGCHROMENLSAKELKIQLPERE
BKXLDHYLPDSRPSDGLGIESALMSGVHGFATJEDDDALMORVEMANSHPAYTNL
SGVALQTESRVTLGAIVAEVNAFNSLPLOVALIOLLAERFENIOSHALVESHK
KISHLACTOSTLEALNPDIPIVSYLSL"
complement (11675..11812)
/locus_tag="VV12380"
/locus_tag="VV12380"
/locus_tag="VV12380"
complement (11675..11812)

Query Match 42.5%; Score 339; DB 1; Length 306147;
Best Local Similarity 66.5%; Pred. No. 1.7e-88;
Matches 486; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 64 TTGTTATTTAGGAGACCATTAATGTCACATCGATGATTTGTCGCTCTTGG 123
DB 5155 TTGATATTAAAGAGTTCCTGATGATTTGATACATGACATGATTTGCGTGAAGTAGGT 5214
QY 124 CCGAATCTATCCAGGTTTATCGTGAAACGCTCTATATCAAAATGGCCAGATTGTGCT 183
DB 5215 GAGAAAGGCTGGAAGTTCGTGATTTAAACGACCAATCTGAGACACACGACATGGC 5274
QY 184 AAATGGCATTCCTGGCGGAGATAGTATGACGAAGATATGACCCGCTCATGCTGAGAA 243
DB 5275 ATGTGGTCTATTCGGAGAGATTGTGTTGAACACGATTTATCCAAAGGAGGACAA 5334
QY 244 CCGTGCATGAGATTGATGAGACGACGACGATTTGTTGGGAAAAAGTCCATCT 303

Db 5335 CCGGCTGACGCTGATTTTGTGATCTGCGGACGCGATCTGTGTGGGAAAAATGATCC 5394
 QY 304 TATCCATTTTATTCAGCGATCGCGTGTGATGCGAACCCCAACCGCATCCGATGGT 363
 Db 5395 TATCCCTACCTACGATGAGTACCGGATGTGTGATGATGACCAAAAGCGCATCTCATGT 5454
 QY 364 TGGAGTGTGATTTTCCCATTAACGCTTATTAACCCGTGGAATGTCAACAAATGAA 423
 Db 5455 TGGAGGTGATCCATGCGCCCATGACACTACTTAACCAACAAATGTGAAATTTCA 5514
 QY 424 GATTTGTGATGACCCCGACGCGCTAATGTTGATCTTATTAATCTTACTCAAGAA 483
 Db 5515 AACTGTGTGTTGTGTGATCAACATTTGCTGTGCGCTTGTAGTGGTGTGCAAGG 5574
 QY 484 GAATGCGCTGCTTTTGTGATGATGCGGCAATTCAGCATGCGTGGCAAAATTAAGC 543
 Db 5575 GAAGTGTGCTTGTGCTTTGACATGCGGATTAATTCGCTGCAATGACCAAGTGGT 5634
 QY 544 GCTGCGTGAATACACATCCGTGTACTATTTTCTTGAAGAAAGATTTTGTAGGCG 603
 Db 5635 GCGGATGGAATACACATCGGTGTGCTTGTGACCTGCAAAAAGATTTCTGTTCT 5694
 QY 604 GATTTATTTGATGCTTACCCCAATTTGCGGTGCAAGTATATGCAATGCAATTAAGC 663
 Db 5695 GACATCATTCACGCTTACCGCAAGTTGCTGTGATGTGACCGCATGACATCAAGCA 5754
 QY 664 CGTTGATCAATACCGGGGTGATCGTCAATACCAATTAATGCGCATCTTGTAAAGC 723
 Db 5755 CGTTATGACACCGGTGTGATCGTACCGCAATTAATGCTTCAACCAAGG 5814
 QY 724 AAAGAGCCAAACCAAGCCGCTTATCGCTTGCAGTCATGAAGTCACTATTTTCA 783
 Db 5815 AAGGAGGCAAAACCGCTCAGGTCTATAGCTTAGCGGATTAACGTAACCTATTTCA 5874
 QY 784 ACTGTTTACG 794
 Db 5875 ACTTGTGCG 5885

RESULT 4
 AP005085/c 303450 bp DNA linear BCT 05-MAR-2003
 LOCUS AP005085
 DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6.
 ACCESSION AP005085 BA000032
 VERSION AP005085.1 GI:28808465
 KEYWORDS
 SOURCE
 ORGANISM
 . Vibrio parahaemolyticus
 Vibrio parahaemolyticus
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 Vibrionaceae; Vibrio.
 REFERENCE
 AUTHORS
 1 Nasu, H., Iida, T., Sugahara, T., Yamachi, Y., Park, K.S., Yokoyama, K.,
 Makino, K., Shinagawa, H. and Honda, T.
 TITLE A filamentous phage associated with recent pandemic Vibrio
 parahaemolyticus O3:K6 strains
 JOURNAL U. Clin. Microbiol. 38 (6), 2156-2161 (2000)
 MEDLINE 20295086
 PUBMED 10834969
 REFERENCE
 AUTHORS
 2 Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,
 Tagemori, K., Tijiima, Y., Neijima, M., Nakano, M., Yamashita, A.,
 Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H.,
 Hattori, M. and Iida, T.
 TITLE Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 distinct from that of V. cholerae
 JOURNAL Lancet 361 (9359), 743-749 (2003)
 MEDLINE 22508454
 PUBMED 12620739
 REFERENCE
 AUTHORS
 3 Ibaes 1 to 303450)
 Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T.,
 Honda, T., Shinagawa, H., Hattori, M. and Iida, T.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome

INFORMATION RESEARCH CENTER, 3-1, YAMADAOKA, SUITA, OSAKA 565-0871,
 JAPAN (E-MAIL: KEN@GEN-INFO.OSAKA-U.AC.JP,
 URL: HTTP://WWW.GEN-INFO.OSAKA-U.AC.JP/, TEL: 81-6-6879-8365,
 FAX: 81-6-6879-2047)
 This clone was isolated from a patient presenting with acute
 gastroenteritis.
 FEATURES
 SOURCE
 COMMENT
 Location/Qualifiers
 1..303450
 /organism="Vibrio parahaemolyticus"
 /mol_type="genomic DNA"
 /strain="O3:K6"
 /sub_strain="RIMD 2210633"
 /db_xref="taxon:670"
 /chromosome="2"
 81..1574
 /gene="VPA0300"
 81..1574
 /gene="VPA0300"
 /note="similar to EMB:CA6187.1 (A1591787) percent
 identity 68 in 487 aa"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical signal peptide protein"
 /protein_id="BAC61643.1"
 /db_xref="GI:28808465"
 /translation="MKLVKPKRLVLTLLASLPISGVAYAAANNESSIVQOQMGN
 WPTQNEKRVQELVYQRAISAYQAPANVIGMDSSSKFGKINVLPIKMDK
 SRLVLPFNADVVMNVLDELKETPLVAAPEKVGMTDFQRTITVGLIGPRA
 RGVLYLLPENVDPVPGKGFETSTNNVFLPFTVMKKGENDPDAVKIABQR
 IYPLMBEKNICPMKFPNAGORINMVPFDGFYMKMGQVDEBPISALSPYLRVL
 ASIGIVGKPEFTVEKOLLEKAVYAPKMTIAORLLGSKDEBNLYKQDROYERAMA
 GATSEWQSYSLIDORAAFFOYVASSAPAMWRTNAGSKYPTFTDSDGLDNGSH
 QYKHLRANPAPLFRNATVLYNVYDGTMTPTSLSSKNGFDEIKNTDSDIDLYGCP
 QLPNGAPESNITIKTIPERDPLTALRLYGTGIEFFDQWPKDDVLKIK"
 complement (1835..2314)
 /gene="VPA0301"
 complement (1835..2314)
 /note="similar to DBJ:BA843758.1 (AP003138) percent
 identity 57 in 158 aa"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="putative peptide methionine sulfoxide reductase"
 /protein_id="BAC61644.1"
 /db_xref="GI:28808467"
 /translation="MKETKIFAGGCLMGVQEFKRLPGVISTEAGRANGKTONTQSEY
 DGYAEVETFEFDSQVSIPLNGVFPETIDPSYINQGDVGKRYTGVYSRUTKLE
 IAKSFATRTDADKIAVEVLPLKNVPSDBEHQDLRTPFPNDYCHPLDLHKYKXK"
 2493..2882
 /gene="VPA0302"
 2493..2882
 /note="similar to GB:AA64344.1 (L16865) percent identity
 52 in 129 aa"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC61645.1"
 /db_xref="GI:28808468"
 /translation="MKIEHVAIWTERLEELKGFYKRYFNAVSNDRKYNPKGSSYFL
 SPESGARLEIMSGEYTCENSHMSVOTGLAFAPALGSEQAVDQITKTLVBDGYORI
 DGPRTGDDGYBSCLDPDGNRIELTV"
 2974..3714
 /gene="VPA0303"
 2974..3714
 /note="similar to GB:AA120622.1 (AE008775) percent
 identity 54 in 243 aa"
 /codon_start=1

```
/evidence=not experimental
/transl_table=11
/product="putative regulatory protein"
/protein_id="BAC61646.1"
/db_xref="GI:28808469"
/translation="MNPQWELIQLVNRKRVQVTELSIDIGSVTLIRQDINPEEQ
GYLKVGAATALQSDIDITRLERFDIKQTLANKADLVAPNEVLIIGGSAALLA
RTLAERGDVITITPSAVIATLIENTSANTIIIGVYQHOGESIVGPTLCLIENTHFS
TATLGDGFRQDGTGTSRDMRADIAEALIAKRRIYVITDSKRGQIIPSSIGKTNE
ISVLTDKAARSDLSQKLGVEVVLG"
/complete="3733. .4089"
/gene="VPA0304"
/complement (3733. .4089)
/notes="similar to GB:AAF96321.1 (AE004376) percent
identity 76 in 117 aa"
/gene="VPA0304"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC61647.1"
/db_xref="GI:28808470"
/translation="MHOEKNTYVERGTENIGATKAPFPHVFGMGFVDYGDYPAAPSG
QGLDGGFYSAEQVQTSNGAALLVYSSDIHATLEKVAFGQIIRPIEPFGGRFH
FTSPSGNEPFAVMSH"
/complete="4319. .5860"
/gene="VPA0305"
/4319. .5860
/gene="VPA0305"
/4319. .5860
/notes="similar to REF:NP_231225.1 (NC_002505) percent
identity 83 in 510 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="catalase"
/protein_id="BAC61648.1"
/db_xref="GI:28808471"
/translation="MNSTHWRATQMSKSFLLITVGLASTLSQAQTLTRDNGAPVGDN
ONSITAGENSVLQDVHLIQLQFARERIPRVVHAGTGAHGFVAGSDPSLTV
SAPTEKGVTPVFPSTVTHSGPSETLDRPGATFTEYEGWMDVGNLNVFV
IRDSIKRPMVHSLKSPVNTNODPNRPFPSHSPATHMLTWYSNIFPGASRTM
DGRVHAIRKINQGDVNYKRFQKSCQCKSLRPRKVTENMGKDPNHLTNLYAITE
RGNYPRKDLTVKVLSPALSKDYNGLDARKWLVNPDKVGWMTLNRLPENFLETE
QSAFAPSNLIPGLEPSBDRLLQGRLEFVADTOLYRLGANLPLPVNRPITSVNNHQN
GLSNNAQLNSGDVNPERSKLNLAEDNOFAVETKLVGTQCKAISKPEDFYQAQVLY
RSNMEQDRSLINLAGLKNKVIDKDIKATMVSFYRADKEVGSRLAEXATNLSQVK
NKAMM"
/5962. .6429
/gene="VPA0306"
/5962. .6429
/notes="similar to GB:AAF94738.1 (AE004235) percent
identity 63 in 152 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="ANKB protein"
/protein_id="BAC61649.1"
/db_xref="GI:28808472"
/translation="MKXVFLSSVLSFSLGSLFSLPASODKLETOYOSLVELFPDAARI
GNBIVETPVSQGPFDORNNESYALMAAYOGKNDCTYRLILDSGANLQDDKQNT
ALMGALIKRISIAKDIYQAECSPLDRNKAAGDLKTFARIYQGSVTLKSLQMK"
/complete="6548. .7201"
/gene="VPA0307"
/6548. .7201
/complement (6548. .7201)
/notes="similar to GB:AAF42465.1 (AE002563) percent
identity 47 in 200 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative pyrazinamidase/nicotinamidase"
/protein_id="BAC61650.1"
```

```
/db_xref="GI:28808473"
/translation="MTYRVNVQPTAVTDVDPKGSFSLGDELVPAGALDVPBLKN
HAKRLKLVSDLHPPKYANDAEITPANKMEPVGLPNDVKNRHCVTGTTGVLIDBL
EPVLVDYQVQKMGDPDHPYGIFFHVDATKGTGAEFKCNKIDIVVGGALDPC
VKSMQALDGEFVIVNLAIRAVLPDITDVSVAEMKEKGLFVNKADIIVERFA"
/complete="7234. .7959"
/gene="VPA0308"
/7234. .7959
/complement (7234. .7959)
/notes="similar to GB:AAF96011.1 (AE004352) percent
identity 60 in 241 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC61651.1"
/db_xref="GI:28808474"
/translation="MIVTVDIIPRLSGCADKGLVLLIKRSPNRPYHGVWALPGF
VFDDLTSEGRPADENFEAARIRCREKTHYRHFSEAFIDDPREDDESLNT
HYALVDNVEQINAGVPGOLKMFLOIINGEETLADHOKTIEKAWOKRASIE
YTVLTLRLDKPEFVADITSAQYFQIDISRMITKRRLISGVLKPTNKVASTKRG
GKPMVYTLNSDEVTFFQNCIRG"
Query Match 35.4%; Score 282.8; DB 1; Length 303450;
Best Local Similarity 63.6%; Pred. No. 6.9e-72;
Matches 431; Conservative 0; Mismatches 247; Indels 0; Gaps 0;
QY 120 TGCGCGGAATCTATNCAGAGTTTACTGCGTGAAGAGCGCTTAATGCCAATGCGGACAGATTG 179
DB 7915 TCGCGATTAAGAGATTGGAAGTGTGCTGATTAAAGCTTCAATCGAATCGGCTTACCA 7856
QY 180 TGGTAATGGCATTTGCGTGGCGGAGATAGTATAGCAAGATATGACCGCTCATGGTGG 239
DB 7855 TGGTGTCGTCATTACCGGCGGCGGTTGCTTTGATTAAGACTTACACAGCGAAGCGG 7796
QY 240 AGAAGCTGTCATGAGATTTTGATGACAGAGACGACGATTTGTGCGCAAAAGTCCA 299
DB 7795 AGCTCGCGGATGATAAACTTTGAAGCGGCGCGCGACGCTTTGCGCGGAAGATTCA 7736
QY 300 TACTTATCTTAATTTCATGACGATCCGCTGTGATGCGCAACCCCAACGCGATCCGAA 359
DB 7735 TACCTACTCTGTCATCTACGACGAGAGGCTTATGAGAGGAGACCTTAAGCGCATCCGGA 7676
QY 360 TGGTTGAGTGTCAATATTTCCCATACGCTTTTAAACCCGTGGAATGTCAACAAAT 419
DB 7675 AGATTGAGCTTAAACATCAACCTTACGATTTGTTGACCGCAACACGCGGCAAT 7616
QY 420 AGAAGATTGGTATGACAGCCCGGCGGCTTAATGGTTGATCTCATCTTACTCAA 479
DB 7615 CAACATGCTGCGCTGCCCGAGTGTCACTCAATGTTCCCATGCAAGCGATCTCAA 7556
QY 480 AGAAGAAATGCGGTGCTTTGATCATGTGCGCAAAATTCAGATGCGTGGCAAAATT 539
DB 7555 TGAAGAAACCTAGCATTTGACCAACAAAACATGCAAAAACATGGCGAAACT 7496
QY 540 ACGCGCTGCGGTGAATAACATCCGCTGATCTATTTCTAGAAAAGAGTTTATG 599
DB 7495 GCGTGTCTGATGATGATCAATAGTGTGCTGTGTGCACTCATTAAGATTTTGGT 7436
QY 600 GCGCGATTTATGATGCTTACGCCAAATTTGGCGTGAAGTTAATGCGATGACCATTA 659
DB 7435 CGCAGACATCATTTCTGATATCAAGATTTGGCATGACATGCGCGCATGACATTA 7376
QY 660 ACGCGCTTATCAATACCGGGGTGATGCTGACATGACATTAATGCGCGCATTTGTA 719
DB 7375 GCGCGTTTGAATCACTAGGTGTGTGAGACCGACCAACAAATCGGTGATCAACAA 7316
QY 720 AGGCAAGAGCCAAACAGGACCATGTTATGCTTTGCCAGTGCATGAGTCACTAATT 779
DB 7315 AGGCAAGCGGTGAAGCAGCATGTGTACAGCTGCAAGCATGAGTGAAGTCTTTT 7256
QY 780 TCAAACTGTTTACGAGG 797
|||||
```


| DB | 7255 | TCAAAACGTGTTGGCGTG | 7238 |
|-----------------------|--|-----------------------|------|
| RESULT 5 | | | |
| LOCUS | AX083744 | 1141 bp | DNA |
| DEFINITION | Sequence 22 from Patent WO011061. | | |
| ACCESSION | AX083744 | | |
| VERSION | AX083744.1 | GI:1185472 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | Synthetic construct | | |
| REFERENCE | 1 | artificial sequences. | |
| AUTHORS | Kunst, L. and Clemens, S. | | |
| TITLE | Regulation of embryonic transcription in plants | | |
| JOURNAL | Patent: WO 011061-A 22 15-FEB-2001; | | |
| FEATURES | UNIVERSITY OF BRITISH COLUMBIA (CA) | | |
| source | Location/Qualifiers | | |
| | 1..1141 | | |
| | /organism="synthetic construct" | | |
| | /mol_type="unassigned DNA" | | |
| | /db_xref="taxon:32630" | | |
| | 1..1141 | | |
| | /note="Consensus sequence of A.t., L.a., and B.n. PAB1 | | |
| | promoters" | | |
| ORIGIN | | | |
| Query Match | 5.3%; Score 42; DB 6; Length 1141; | | |
| Best Local Similarity | 8.3%; Pred. No. 0.62; Mismatches 357; Indels 3; Gaps 1; | | |
| Matches | 59; Conservative 295; Mismatches 357; Indels 3; Gaps 1; | | |
| QY | 8 TCACAAGCTTATTACATTGACCATCATGCACTTAACCTAAATAAGCCGCTGT | 67 | |
| DB | YCKRRWYNNKSRMKGWKKKRYBCANNBSRYHARRKDKYVBMTHNNKMGKTGR | 124 | |
| QY | 68 TTATTAGGAAGCACTTATGATGTGCACTATGCAATATATTTGCTGCGTCTGGCCGA | 127 | |
| DB | HRWYRMRMBTVDHVVYTAANNAAVTTTCMDKDKRTRMWWKNNATGMDDDTKYMW | 184 | |
| QY | 128 AATCATTCAGGTTTACTGTCAGAGACACGCTTAATCAATTCGGCAGATGTGTAAT | 187 | |
| DB | NNNGGBVTWYVRYKTRDWSBKRNWYMBWKNNSYVTTYWWMDCRKRYRWRVT | 244 | |
| QY | 188 GGGCATTCGCTGGCGGAGTAGTGTATGACGAAGATATACCGCTCATGCTGGAGAACTG | 247 | |
| DB | RGRRWYVVAEBTHARRNNGWBYBAMVRYWYNNNNNNNAKAKCKRKYWGWRABVNS | 304 | |
| QY | 248 TCGATGAAGATTTGATGACGAGACACACGATATTTGTCGGCAAAAAGTCATCTATC | 307 | |
| DB | TCTTW--KSKTTTKRISCCANNRCAGANRDKHKWKSAAAGYVNNNNNNNNNTYKAR | 361 | |
| QY | 308 CTAATTATTCAGCGATCGCTGCTGTTGAGCAACCCCAACGCGATCGAATGGTGA | 367 | |
| DB | HBAWMDWYWSAMKKEHANAAYSRKKTWBYKRTMYVNNNGTTWKKMYAMWYKMDW | 421 | |
| QY | 368 GTGTCAGATTTCCATAGACGCTTATTAACCCGCTGGAANGTCAAAAGATAGAGATT | 427 | |
| DB | BGTNNNNNGRATTYYGWTNKKMYTYWKALNNCKRMRMDKTCNNNTTWKCKTYNN | 481 | |
| QY | 428 TTGGTATGACCCCGAGCGCGCTAATGGTTGATCTTTCATACCTTACTCAAGAAGA | 487 | |
| DB | CYWSMTNKGSRBAALAYTYWYMMRRYAHANNNNNDWYWKACTWYKVCBSKMNMYA | 541 | |
| QY | 488 TGGCGCTGCTTTGATGATGTCGGCAAAATTCAGATGCGTGGCAAAAATTACGCGCTG | 547 | |
| DB | AWYKSSWYYSRYRMRKTNNSMKRSDTRMGRRNYPARABHYGYKNTFRWMSHTWB | 601 | |
| QY | 548 CGGTTGATACACATCCGCTGCTATTTATTTAGAAAAAGATTTTATGCGCGATA | 607 | |
| DB | HBRAAAHYMMBYBAKCHCKMAYKATKATGAGGSSNNNNNNNNNNNNNNNNNNNNNN | 661 | |
| QY | 608 TTATTGATGCTACGCCAAATTTGGCGTGAAGTTATGCAATGCAATTAAGCCGCT | 667 | |

| | | | | |
|------------|--|------------|---|-----|
| | Db | 662 | YVLAASRVTAAATAACATGTTTAAANNATYTTHANNWGCNNATDTRRTWKNNNNNAAGTWK | 721 |
| Oy | LOCUS | 668 | TGATCATPACCGGGGTGANTCTGTCAGTACCATAAATATGCCGCACTCTTGTAAG | 721 |
| Db | DEFINITION | 722 | NNNNNNAKNAASAATGYAAAAVKAKCKEWRNANWAMGWHADAAAATTDRKNNG | 775 |
| RESULT 6 | AC138591 | | | |
| LOCUS | AC138591 | | | |
| DEFINITION | Mus musculus chromosome 3 clone RP24-224A19 map 3, WORKING DRAFT SEQUENCE, 7 unordered pieces. | | | |
| ACCESSION | AC138591 | | | |
| VERSION | AC138591.3 | | | |
| KEYWORDS | HTG, HTGS PHASEI, HTGS DRAFT. | | | |
| SOURCE | Mus musculus (house mouse) | | | |
| ORGANISM | Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| REFERENCE | Bairren, B., Nusbaum, C. and Lander, E. Mus musculus chromosome 3, clone RP24-224A19 Unpublished 2 (bases 1 to 146963) | DNA linear | HTG 23-MAR-2003 | |
| AUTHORS | Bairren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barma, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgaglier, B., Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Haez, N., Hagos, B., Harton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menues, L., Mihova, T., Mienga, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severly, P., Smith, C., Spencer, B., Stange-Thomann, U., Stojanovic, N., Talamae, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vasilev, H., Viel, R., Vo, A., Wildson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M. | | | |
| JOURNAL | Submitted (10-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA | | | |
| REFERENCE | 3 (bases 1 to 146963) | | | |
| AUTHORS | Bairren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barma, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgaglier, B., Canarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearlano, K., Diaz, J.S., Dodge, S., Doyle, K., Dorris, L., Erickson, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Haez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mahbit, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Menues, L., Mihova, T., Mienga, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severly, P., Smith, C., Spencer, B., Stange-Thomann, U., Stojanovic, N., Strubbs, M., Tamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vasilev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M. | | | |
| JOURNAL | Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA | | | |
| COMMENT | On Mar 23, 2003 this sequence version replaced gi:28275009. All repeats were identified using RepeatMasker: Smrt, A.F.A. & Green, P. (1996-1997) | | | |

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 128955

Center clone name: 224.A.19

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145843 bases at least Q30

Consensus quality: 146107 bases at least Q30

Consensus quality: 146232 bases at least Q20

Insert size: 146000; agarose-fp

Insert size: 146363; sum-of-contigs

Quality coverage: 15.5 in Q20 bases; agarose-fp

Quality coverage: 15.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2163: contig of 2163 bp in length
* 2164 2263: gap of 100 bp
* 2264 5396: contig of 3133 bp in length
* 5397 5496: gap of 100 bp
* 5497 11945: contig of 6449 bp in length
* 11946 12045: gap of 100 bp
* 12046 19688: contig of 7633 bp in length
* 19689 19788: gap of 100 bp
* 19789 39404: contig of 19616 bp in length
* 39405 140276: gap of 100 bp
* 140277 140376: contig of 100772 bp in length
* 140377 146963: contig of 6587 bp in length.

FEATURES

source

1.146963
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="3"
/map="3"

/clone="RP24-224A19"

/clone_lib="RP24 Male Mouse BAC"

1.2163

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

2264.5396

/note="assembly_fragment"

5497.11945

/note="assembly_fragment"

12046.19688

/note="assembly_fragment"

19789.39404

/note="assembly_fragment"

39505.140276

/note="assembly_fragment"

140377.146963

/note="assembly_fragment"

clone_end:T7

vector_side:right"

ORIGIN

Query Match 5.0%; Score 40.2; DB 2; Length 146963;
Best Local Similarity 52.7%; Pred. No. 3.5;

Matches 87; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 411 CAACCAATGAGATTTTGTATCGACCCGCGCGTAAATGTTGATCTTATAC 470

Db 123522 CAGATGAAAGAAAGATTTTGGCGCTATCCAACTGCTGATCTTAAGAAACAATA 123561

QY 471 TTACTCAAGAAAGAAATGCGCGCTTTTGATCATGTCGCGCAATTCAGCATCGTG 530

Db 123582 ATTGCTAATTAGAGAGAGAAAGGCTTTCTCTGTGTCAGAGAAATTAAGTCGAA 123641

QY 531 GCMAAATTCGCGCTGCGGTTGAATACATCCGTGTAATTT 575

Db 123642 GAGAAATTCACACCTCATAGGAATCATCATTCATGAGATT 123686

RESULT 7

AF146758

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNALS

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

source

gene

mRNA

CDS

FEATURES

Query Match 5.0%; Score 40; DB 3; Length 2067;
Best Local Similarity 51.7%; Pred. No. 2.6;

Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 452 ATGGTTGATCTTCACTTACTCAAGAGAAATGCGCTGCTTTGATCATGTCG 511

Db 1819 AATGGAAGATTTCACTTACTTAAAGAAATGCTGACCTTTATTTGATTC 1878

QY 512 CCGAATTCAGATGCGTGGCAAAATATACCGCGTGGTGAATACATCCGTGGAC 571
 DB 1879 GAGAGGATTTTCGTAATGAATGACAGAGCTGAGTTGCAAGAGAAATGCTACTCT 1938
 QY 572 TATTTTCATTGAAAAAGATTTTATGTCGGGATTAATGATCCCTACGCCAA 627
 DB 1939 TTTTATTAAGAAAAATTTTATGCGTGGCTTATGCTTTTTCGACACCAAA 1994

RESULT 8
 AC012853/c 16237 bp DNA linear HTG 03-NOV-1999
 LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 DEFINITION AC012853
 ACCESSION AC012853
 VERSION AC012853.1 GI:6223468
 KEYWORDS HTG; HTGS PHASE2.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 16237)
 AUTHORS Adams, M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10210833 by the submitter. For further information on this sequence you may e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..16237
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

ORIGIN
 Query Match 4.9%; Score 39.2; DB 2; Length 16237;
 Best Local Similarity 47.5%; Pred. No. 5.5; Mismatches 163; Indels 2; Gaps 1;
 Matches 149; Conservative 0;

QY 330 GGTGATGCGAAGCCCAAGCGATCCGATGTTGAGTGTCAATTTCCCATTCGC 389
 DB 2498 GGTAAATATATCATTAACGGAATGATGGGTGAATCTAAATATTAATGTTAATCAT 2439
 QY 390 TTTATTAACCCGTGATGTCMAACAATAGAAATTTTGTATTCGACCCCGCGGC 449
 DB 2438 ACTTACTACTGTTTAAATGCAAGTAACTGAAATTTAAGTATTCAAATATTGGC 2379
 QY 450 TAAATGTTGATCTTCACTTACTTCAAGAAAGAAATGCGCGTGTTCATGATGT 509
 DB 2378 TAAATGATATATATCTTTTAAACGCCAGAAAGAAAGCTATATCAGTTCAATGT 2319
 QY 510 CGCGCAATTAACGATGCGTGGCAAAATTTACGCGCTGGCTTAATACATCGTGT 569
 DB 2318 GAATTAACCAAGGCACTAACTGAAATCCGCAATATCTACCTTAACGATCTTAT 2259
 QY 570 ACTATTTTCATTA--GAAAAAGATTTTATGCGGATTAATGATGCTACGCCAA 627
 DB 2258 CACATGCAATGTATGAATTAATTTGACGCGGAATTTTGCAATATCAACGTAGA 2199
 QY 628 TTTGCGCTGGAAGT 641
 DB 2198 TTTGCGACTCGGACT 2185

RESULT 9
 AC012650/c 74085 bp DNA linear INV 12-MAY-2001
 LOCUS AC012650

DEFINITION Drosophila melanogaster, chromosome X, region 19B-19E, BAC clone BACN24N09, complete sequence.
 ACCESSION AC012650
 VERSION AC012650.8 GI:14028960
 KEYWORDS HTG.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 74085)
 AUTHORS Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Bonzon, J., Beeson, K.V., Busan, D.A., Carlson, J.W., Center, A., Champagne, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresner, D., Farfan, D., Ferreira, S., Frishe, E., Galle, R.F., Gary, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ilegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A., Mincosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Paclob, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Pfenner, S., Pfenner, M., Strong, R., Svitzkas, R., Tector, C., Williams, S.M., Zaveri, U.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 TITLE Sequencing of Drosophila chromosome X, region 19E-19B
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 74085)
 AUTHORS Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazer, R.G., Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karia, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomoclan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S., Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Snir, E., Svitzkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
 TITLE Direct Submission
 JOURNAL Submitted (02-NOV-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
 COMMENT On May 12, 2001 this sequence version replaced gi:5958011.
 Sequence submitted by: Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdgs@fruitfly.berkeley.edu.
 FEATURES
 source
 1..74085
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="hy; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="X"
 /map="19E-19E"
 /clone="BACN24N09 (D1133)"
 /clone_11b="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pbelBac11)"

ORIGIN
 Query Match 4.9%; Score 39.2; DB 3; Length 74085;
 Best Local Similarity 47.5%; Pred. No. 6.5; Mismatches 163; Indels 2; Gaps 1;
 Matches 149; Conservative 0;

QY 330 GGTGATGCGAAGCCCAAGCGATCCGATGTTGAGTGTCAATTTCCCATTCGC 389
 DB 67158 GGTAAATATATCATTAACGGAATGATGGGTGAATCTAAATATTAATGTTAATCAT 67099
 QY 390 TTTATTAACCCGTGATGTCMAACAATAGAAATTTTGTATTCGACCCCGCGGC 449

DB 67098 ACTTACTACTGTTTATTTATTCAGAGTAATGTAATTTAAGTATTCATTAATTTGGC 67039
QY 450 TAATTGGTTGATCTTCACTTACTCAAGAGAAATGCCGCTGTTGATCATGT 509
DB 67038 TAAATAGTATATATCTTTTAAACCCGCAAGAAAGAAAGCCCTATATCATGTTCTAAATGT 66979
QY 510 CGGCAAAATTCAGCATGCGGCAAAATTAAGCCGCTGCTGATACATCCGCTGT 569
DB 66978 GAATCAAAACAGCAATTAAGTATGTAATGTAAGCGGCAATTTGATATCAACGTCTATT 66919
QY 570 ACTATTTTCATTA--GAAAAGAGTTTATGAGCGGATATTTATGATGCTTACGCCCAA 627
DB 66918 CACATGCGAATGATATTAATTAATTTAGACGCGCAATTTGATATCAACGTAGA 66859
QY 628 TTTCGCGTGGAGT 641
DB 66858 TTTCGACTCGGACT 66845

RESULT 10
AC011761 108350 bp DNA linear INV 06-SEP-2001
LOCUS Drosophila melanogaster, chromosome X, region 19B-19E, BAC clone
DEFINITION BACH50G05, complete sequence.
AC011761 GI:15451480
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 108350)
Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, J., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Buesam, D.A.,
Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Fairman, D.,
Ferreira, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jaitai, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McInosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuro, J.,
Paclob, D., Paragae, V., Park, S., Patel, S., Pfeiffer, B.,
Phonemavong, S., Piltman, G.S., Puri, V., Richards, S., Scheeler, F.,
Shapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome X, region 19E-19E
Unpublished
2 (bases 1 to 108350)

REFERENCE
AUTHORS
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blasej, R.G.,
Butenchoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Katta, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S.,
Pfeiffer, B., Poon, L., Segura, A., Sechi, H., Smit, B.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Laboratory (14-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 6, 2001 this sequence version replaced gi:14280137.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

COMMENT
TITLE
JOURNAL
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES
source

Location/Qualifiers
1..108350
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="19E-19E"
/clone="BACH50G05 (D1134)"
/clone_11b="RPCT-98 (Rosewall Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBelBAC11)"

Query Match 4.9%; Score 39.2; DB 3; Length 108350;
Best Local Similarity 47.5%; Pred. No. 6.8; Index 2; Gaps 1;
Matches 149; Conservative 0; Mismatches 163; Indels 2; Gaps 1;

QY 330 GGTGATGCGAACCAGCCGATCGAATGTTGAGTGTGATTTCCATTACGC 389
DB 104731 GGTAAATTTATCATTTAAACGCAATGATGAGTGAATTAATTTACTGTAAAT 104790
QY 390 TTTTAAACCCGCGAATGCAAAATAGAAATTTGATATGACCCGAGCGC 449
DB 104791 ACTTACTACTGTTTATTTATGCAAGTAACTGAATTAATTAATTTACTGTAAAT 104850
QY 450 TAATTGGTTGATCTTCACTTACTCAAGAGAAATGCCGCTGTTGATCATGT 509
DB 104851 TAAATAGTATATATCTTTTAAACGCAAGAAAGAAAGCCCTATATCAATTTAATGT 104910
QY 510 CGGCAAAATTCAGATGCGGCAAAATTAATGCGCGCTGTTGATACATCCGTGT 569
DB 104911 GAATCAAAACAGCATTAATTAATGTAATGTAAGCGCAATTAATTAATTTCTATT 104970
QY 570 ACTATTTTCATTA--GAAAAGAGTTTATGAGCGGATATTTATGATGCTTACGCCCAA 627
DB 104971 CACATGCGAATGATATTAATTAATTTAGACGCGCAATTTGATATCAACGTAGA 105030
QY 628 TTTCGCGTGGAGT 641
DB 105031 TTTCGACTCGGACT 105044

RESULT 11

AE003571/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

AE003571 302540 bp DNA linear INV 14-FEB-2003
Drosophila melanogaster chromosome X section 68 of 74 of the
complete sequence.
AE003571 AE002620 AE014298
AE003571.3 GI:22832655
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 302540)
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Galle, R.F.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutcliffe, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brudon, R.C., Rogers, Y.H., Blasej, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Holt, G., Nelson, C.R., Gabor, G.L.,
Abell, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Basely, E.M.,
Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,
Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolintsov, S.,
Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brotler, P.,
Butler, K.C., Butts, D.A., Butler, H., Cadieu, E., Center, A.,
Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,
Davies, P., de Pablo, B., Delcher, A., Deng, Z., Drenth, A.D., Dew, I.,
Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S.,
Dunkov, B.C., Dunn, P., Durkin, K.J., Evangelista, C.C., Ferraz, C.,
Ferreira, S., Fleischmann, M., Foster, C., Gabrielian, A.E., Garg, N.S.,

Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Hejman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Krenitsen,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Maitel,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusken,D.R., Pacleb,J.M., Palazolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reiner,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstein,G.M., Weisenbach,J., Williams,S.M., Woodgett, Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,U., Yen,R.F., Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

20196006

10731132

2 (bases 1 to 302540)

Celniker,S.E., Adams,M.D., Kronmiller,B., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., Banson,J., An,H., Baldwin,D., Bantz,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Chame,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreznak,D., Fafian,D., Fertler,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nuro,J., Pacleb,J., Paragis,V., Park,S., Patel,S., Pfeiffer,B., Phoonanang,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svitskas,R., Tector,C., Tyler,D., Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.

Sequencing of *Drosophila melanogaster* genome
 Unpublished

3 (bases 1 to 302540)

Mirza,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L., Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochuk,S.E., Smith,C.D., Tupy,J.U., Bergman,C.M., Berman,B.P., Carlson,J.W., Celniker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de Grey,A.D.N.J., Harris,N.L., Kronmiller,B., Marshall,B., Milburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E., Shu,S., Smutniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M., Rubin,G.M., Mungall,C.J. and Lewis,S.E.

Annotation of *Drosophila melanogaster* genome
 Unpublished

4 (bases 1 to 302540)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.

Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

5 (bases 1 to 302540)

REFERENCE

AUTHORS

CONSTRM

FLYBASE

Direct Submission
 Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA

6 (bases 1 to 302540)

REFERENCE

AUTHORS

CONSTRM

FLYBASE

Direct Submission
 Submitted (23-JAN-2003) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA

COMMENT

FEATURES

SOURCE

1. 302540
 Location/Qualifiers
 /organism="Drosophila melanogaster"

/mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /chromosome="X"
 /note="genotype: y[1], cn[1] bw[1] sp[1], Rh6[1]"
 8252. .9439
 /locus_tag="CG1631"
 /note="Last curated on Wed Jul 03 12:02:51 PDT 2002"
 /map="19c3-19c3"
 /db_xref="FLYBASE:FBgn0031101"
 /join(8252. .8476,8536. .8972)
 /locus_tag="CG1631"
 /product="CG1631-PA"
 /db_xref="FLYBASE:FBgn0031101"
 /join(8362. .8476,8536. .8972)
 /locus_tag="CG1631"
 /note="CG1631 gene product"
 /codon_start=1
 /product="CG1631-PA"
 /protein_id="AAF50920.1"
 /db_xref="GI:7295610"
 /translation="MADSPANRDRDQINADVSDISTSRSELTSDYTSDESDCTVADPTSPQSMVSLASDSPLSEQNFCEITSLAHETLCCEGDSGSEVSDG ELDDEARBYDMFGESSEFGLEPYRLMSDEBDNSNEVVGSNPVGETIFPLEKDK DOOSEKSKSKSEKIFQWELFE"
 /complement(47533. .48708)
 /locus_tag="CG15462"
 /note="Last curated on Wed Jul 03 14:50:13 PDT 2002"
 /map="19c4-19c4"
 /db_xref="FLYBASE:FBgn0031102"
 /complement(47533. .47774,48345. .48708)
 /locus_tag="CG15462"
 /product="CG15462-PA"
 /db_xref="FLYBASE:FBgn0031102"
 /complement(47533. .47774,48345. .48708)
 /locus_tag="CG15462"
 /product="CG15462-PA"
 /db_xref="FLYBASE:FBgn0031102"
 /complement(47533. .47774,48345. .48708)
 /locus_tag="CG15462"
 /note="CG15462 gene product"
 /codon_start=1
 /product="CG15462-PA"
 /protein_id="AAF50918.1"
 /db_xref="GI:7295608"
 /translation="MNTAVCYRACACGEGDSVGVQFTGADSESGREMAATLSLSEKFSISLRKQFVDSRQKQVIEKLGQVAVCRFRHSIIDPRDPAITFDA VMLDFQLACVAVCWEMKLRRLHAAHTRVPLPADPTDIDVNVNVNVNV NVNGGYGVADVADVEWDEARMDMDMDVLDKAFHCGMCON"
 /complement(56521. .57280)
 /locus_tag="CG12679"
 /note="Last curated on Wed Jul 03 14:49:36 PDT 2002"
 /map="19c4-19c5"
 /db_xref="FLYBASE:FBgn0031103"
 /complement(56521. .57280)
 /locus_tag="CG12679"
 /product="CG12679-PA"
 /db_xref="FLYBASE:FBgn0031103"
 /complement(56578. .57150)
 /locus_tag="CG12679"
 /note="CG12679 gene product"
 /codon_start=1
 /product="CG12679-PA"
 /protein_id="AAF50917.1"
 /db_xref="GI:7295607"
 /translation="MNVGALVNLIIILACTYMTSLRPADHPYAFLAASPLVGV LGVRSFANPDECGRTFMISATLIVLPLANNIFVYSDSGVALVHGSLTFLFY DMIGKISDWDSTETLMDLALNIGISTLYALAKDGNHIFGVAAATLALRYGAIV DSCNGSLNVVETLTGTLALMTALTEG"
 58131. .59536
 /locus_tag="TBE19654"
 /map="19c5-19c5"
 58131. .59536
 /translation="H-element {}182"
 /db_xref="FLYBASE:FBti10019654"

```

gene      62674..63120
          /locus_tag="TE19655"
          /map="19C5-19C5"
repeat_region 62674..63120
          /transposon="3S18(183"
          /db_xref="FLYBASE:FB10019655"
          /complement(71194..71526)
          /locus_tag="CG15460"
          /note="Last curated on Med Uni 03 14:48:49 PDT 2002"
          /map="19C5-19C5"
          /db_xref="FLYBASE:FBgn0031104"
          /complement(71194..71526)
          /locus_tag="CG15460-PA"
          /db_xref="FLYBASE:FBgn0031104"
          /complement(71194..71526)
          /locus_tag="CG15460"
          /note="CG15460 gene product"
          /codon_start=1
          /product="CG15460-PA"
          /protein_id="FAF50916.1"
          /db_xref="GI:7295606"
          /db_xref="FLYBASE:FBgn0031104"
          /translation="MSQKPKPNNPNESHSTFLNIPBELQPTRGKPMQNGKST
          MKGLPSNTVNSAQTASVNDNLQASVIALMAGMDSLDMEQPNRSSRSHERLINE
          LFFSNML"

Query Match      4.9%; Score 39.2; DB 3; Length 302540;
Best Local Similarity 47.5%; Pred. No. 7.6;
Matches 149; Conservative 0; Mismatches 163; Indels 2; Gaps 1;

QY      330 GGTGATGGCAACCCCAACGCGATCCGAATGGTGGAGTGTCACTATTTCCCAATACGC 389
DB      302342 GGTAAATAATATCATTAAGCGAATGGTGAATGGAATCTAAATATTAATCTGTTAAACAT 302283
QY      390 TTTATTAACCCCTGGAATGTCAACAATAGAAAGTTTGTGATCGACCCGAGCGCGC 449
DB      302282 ACTTACTTAATTTTAAATGCAAGTAACTGAAATTTAAAGTAAATTAATTTGGC 302223
QY      450 TAATGGTTGATCTTCACTTACTTACTCAAGAAGAAATGCGCTGGTTTGATCATGT 509
DB      302222 TAATAGTATATATCTTTTAAAGCGCAAGAAAGAAAGCCATATCAATCTAAATGT 302163
QY      510 CGGCGCAATTCAGATGCGTGGCAAAATTAACCGCGTGGTGAATACATCCGTGCT 569
DB      302162 GAATCAAAACGCGCAATAAACTGAATACCGCAATTAATCCGTAAACAGATTCCTAT 302103
QY      570 ACTATTTTCAATTA--GAAAAGAGTTTATAGTGGCGGATTTATGATGCTACGCAAA 627
DB      302102 CACATGCGAATGATGATTAATTAATTTGTAGCGCGCAATTTGTGATATCAAGCTAGA 302043
QY      628 TTTGGCGTCGAAGT 641
DB      302042 TTCGACTCGGACT 302029

RESULT 12
AC135793/ 128683 bp DNA linear HTG 04-NOV-2003
LOCUS      Oryza sativa (japonica cultivar-group) chromosome 11 clone
DEFINITION OSJNB0017P15, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC135793
VERSION    AC135793.3 GI:38153795
KEYWORDS   HTG; HTGS PHASE2; HTGS_ACTIVEPIN.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 128683)
            Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Gamberger, K., Jones, K.M.,
            Overton, II, U., Tselirin, T., Kim, M., Beta, V., Jin, S., Padrosh, D.W.,
            Tallon, L., Koo, H., Ziemann, V., Hsieh, J., Blunt, S., Vanaken, S.,

```

```

TITLE      Riedmuller, S.B., Uterbach, T., Feldl, Yum, T., Yang, Q., Haas, B.,
            Sub, B., Peterson, J., Quackenbush, J., White, O., Salzberg, S. and
            Fraser, C.
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 128683)
AUTHORS    Buell, R.
TITLE      Direct Submission
JOURNAL    Submitted (22-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
            3 (bases 1 to 128683)
            Buell, R.
REFERENCE   Direct Submission
AUTHORS    Submitted (04-NOV-2003) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
            On Nov 4, 2003 this sequence version replaced gi:24431531.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            1 128683: contig of 128683 bp in length.
            Location/Qualifiers
              1..128683
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="genomic DNA"
                /cultivar="Nipponbare"
                /db_xref="taxon:39947"
                /chromosome="11"
                /clone="OSJNB0017P15"
                /note="japonica cultivar-group"

ORIGIN
Query Match      4.9%; Score 39; DB 2; Length 128683;
Best Local Similarity 11.9%; Pred. No. 7.9;
Matches 48; Conservative 170; Mismatches 185; Indels 0; Gaps 0;

QY      384 TTACGCTTATTAACCGCGAATGCAACAAATGAAGATTTGTGATTCGACCCGA 443
DB      50864 TKATCTTKRKTWTWTCTMTKCKRCGCMWMAWYMYTYTWTWTWCCATACMMW 50805
QY      444 GCGGCTAATTTGTTGATCTTCACTTACTCAAGAAGAAATGCGGCTTTGA 503
DB      50804 WMMMAAMMMWMMWMMWCMYKAYMMATWTATASACTTGGAWYTCYTTGGRKYRTCT 50745
QY      504 TCATGTGCGCAATTAAGCATGCGTGGCAAAATTAACCGCGTGGTGAATACATC 563
DB      50744 KMYKSTTTGCAWMTAKGMYGMSKTWTGSCYKRYRMYCYMSMKRKKSTAYR 50685
QY      564 CGTGATCTATTTTCAATTAAGAAAGATTTTATAGTGGCGAATTTATGATGCTACGC 623
DB      50684 YMYSAWYSWTSYKKAAYRRGSCMAWYKMAAAMWYTYMYMKRGRAMMYCRM 50625
QY      624 CAATTTGGCGTCGAAGTTAATGCATGACCATTAAGCGCGCTTGATCAATACGGGGT 683
DB      50624 WMAAMWYRCMTKRTGTGASWKMSPAAAKYMAWYCRMRKRYKAYSTRMKYKMAWMA 50565
QY      684 GATGTCAGTACCAATTAAGTGGCGCATCTTTGAAGGCAAGAGCAACGACGAC 743
DB      50564 TWYHGRARWYMYMKMBWTTMMARAKKATRAAASSMAKRRRTYMSASRRKM 50505
QY      744 CGTTATGCTTCCAGTACGATGAAGTACCTATTTTCAAC 786
DB      50504 MWMGMSRRYWMSWMMRKRKYKAYKACAYEMWASY 50462

RESULT 13
AC119922

```

LOCUS AC119922 168422 bp DNA linear HTG 12-MAR-2003
 DEFINITION Mus musculus clone RP24-252P2, WORKING DRAFT SEQUENCE, 6 unordered
 pieces.
 AC119922
 AC119922 3 GI:26927735
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 168422)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP24-252P2
 Unpublished
 2 (bases 1 to 168422)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Bougialter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
 Charaz, B., Choepel, Y., Colangelo, N., Collins, S., Collins, A.,
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamet, A., Karatas, A., Kelle, C., Lacroque, K., Lamaras, R.,
 Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Margis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunhphang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retts, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Roedel, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 168422)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, S., Arauchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Bougialter, B., Camarata, J., Chang, J., Choepel, Y.,
 Colangelo, A., Cooke, P., Cooke, P., Corum, B., Deatellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, J., Erickson, J., Fato, S.,
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamet, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunhphang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retts, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talama, J., Teefaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 12, 2003 this sequence version replaced gi:26201661.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center Project name: 125309
 Center Clone name: 252.P.2
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 167327 bases at least Q40
 Consensus quality: 167693 bases at least Q30
 Consensus quality: 167845 bases at least Q20
 Insert size: 163000; agarose-ffp
 Insert size: 167922; sum-of-ctnigs
 Quality coverage: 12.7 in Q20 bases; agarose-ffp
 Quality coverage: 12.3 in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 11641: contig of 11641 bp in length
 * 11642 11741: gap of 100 bp
 * 11742 14593: contig of 2852 bp in length
 * 14594 14693: gap of 100 bp
 * 14694 78832: contig of 64139 bp in length
 * 78833 78932: gap of 100 bp
 * 78933 115186: contig of 36254 bp in length
 * 115187 115286: gap of 100 bp
 * 115287 115317: contig of 36031 bp in length
 * 115318 151417: gap of 100 bp
 * 151418 151418: contig of 17005 bp in length.
 * 151418 168422: contig of 17005 bp in length.
 Location/Qualifiers
 1. 168422
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP24-252P2"
 /clone_1lb="RP24-252P2"
 /clone_1lb="RP24-252P2"
 1. 11641
 /note="assembly_fragment"
 clone_end:8P6
 vector_side:left
 misc_feature
 /note="assembly_fragment"
 11742..14593
 misc_feature
 /note="assembly_fragment"
 14694..78832
 misc_feature
 /note="assembly_fragment"
 78933..115186
 misc_feature
 /note="assembly_fragment"
 115287..115317
 misc_feature
 /note="assembly_fragment"
 151418..168422
 /note="assembly_fragment"
 clone_end:17
 vector_side:right
 ORIGIN
 Query Match 4.9%; Score 39; DB 2; Length 168422;
 Best Local Similarity 46.2%; Pred. No. 8.1;
 Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
 QY 448 GCTAATGTTGATGCTTCACTTCACTCAAGAAAGAAATGCGCTGCTTGAATCAAT 507
 Db 136013 GTTGTGTTGTTTCTTCACTATATATAGAACACATTTTCCCTTTCTTATTAAT 138072
 QY 508 GTGCGCAATTCAGCATCGTGGCAAAATTCACGCGCTGCGGTGAATACACATCCG 567
 Db 138073 ATATGCTGAAAAACAGTCTTATCAAGAAATGTGTGACGAGATTAACATCAACAA 138132
 QY 568 GTACTATTTTCATATAGAAAAAGATTTTAGTGGGGGATATATATGATGCTTACGCCA 627

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS

1
Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kikuchi, H., Murakawa, K., Kaneshiro, K., Takahashi-Fujii, A.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T.
MEDO human cDNA sequencing project

TITLE Unpublished

2 (bases 1 to 2669)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT MEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
Location/Qualifiers

1..2669
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PERIC1000147"
/cissue_type="pericardium"
/clone_lib="PERIC1"
/note="Cloning vector: PME18SFL3"

ORIGIN

Query Match 4.9%; Score 38.8; DB 9; Length 2669;
Best Local Similarity 53.2%; Pred. No. 6;
Matches 82; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 241 GACCTGTCGATGAGATTTTGTATGACGAGACGACGATTTGTGGCAAAAAGTCCAT 300
DB 268 GGACCTTTCATTCATTTTAAAGCTGTGAGATATTTGTTTCATTAGAAAGTCGTG 209
QY 301 ACTTATCCATTTTATCAGGATCCGCTGGTGTATGSCAACCCCAAGCGATCCGAT 360
DB 208 GCTTTTCAAGTGAACATAGTCTCCCGCTTTTGTGTGGTGCCTTAATGATGATGACT 149
QY 361 GGTGAGTGTGATTTCCCATTAAGCTTTAT 394
DB 148 TTTTCAGAGACTTTCATTCTGTACATGTGAT 115

Search completed: April 30, 2004, 15:29:03
Job time : 3230.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 30, 2004, 06:19:01 / Search time 326.63 Seconds
(without alignments)
10378.910 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798

Sequence: 1 atgttcgtcacaaagtccta.....ttcaacctgttcacaggt 798

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_28Jan04:*

1: Geneseq19808:*\n2: Geneseq19908:*\n3: Geneseq20008:*\n4: Geneseq20018:*\n5: Geneseq20028:*\n6: Geneseq20038:*\n7: Geneseq20048:*\n8: Geneseq20058:*\n9: Geneseq20068:*\n10: Geneseq20078:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 405.8 | 50.9 | 717 | 7 | ACA53528 Prokaryot |
| 2 | 38.2 | 4.8 | 6065 | 8 | AAD57245 Human CGD |
| 3 | 38 | 4.8 | 2000 | 7 | ADA71938 Rice gene |
| 4 | 37.6 | 4.7 | 2000 | 7 | ADA71938 Rice gene |
| 5 | 36.8 | 4.6 | 438 | 5 | ABA18862 Human ner |
| 6 | 36.8 | 4.6 | 438 | 5 | ABA18862 Human ner |
| 7 | 36.8 | 4.6 | 438 | 5 | ABA18862 Human ner |
| 8 | 36 | 4.5 | 73465 | 6 | ABQ88161 Human oet |
| 9 | 35.6 | 4.5 | 552 | 9 | ADC90659 E. faeciu |
| 10 | 35.4 | 4.4 | 1119 | 9 | ADB76952 Mouse CLC |
| 11 | 35.2 | 4.4 | 4368 | 2 | AAQ49902 Glutamic |
| 12 | 35.2 | 4.4 | 4512 | 6 | AB199832 Mouse isc |
| 13 | 34 | 4.3 | 110000 | 6 | ABQ67196-5 of |
| 14 | 34 | 4.3 | 110000 | 6 | ABQ67196-5 of |
| 15 | 34 | 4.3 | 110000 | 6 | ABQ67196-5 of |
| 16 | 33.8 | 4.2 | 1974 | 6 | AA146803 H influen |
| 17 | 33.6 | 4.2 | 2628 | 2 | AA224773 Human sol |
| 18 | 33.6 | 4.2 | 2628 | 6 | ABK49992 Degenerat |
| 19 | 33.6 | 4.2 | 110000 | 6 | ABK49992 Degenerat |
| 20 | 33.4 | 4.2 | 1776 | 7 | ADA70920 Rice gene |
| 21 | 33.2 | 4.2 | 223 | 7 | AA537288 Novel hum |
| 22 | 33.2 | 4.2 | 2622 | 7 | ACA28161 Prokaryot |
| 23 | 33.2 | 4.2 | 82952 | 6 | ABN85766 Arabidops |

| | | | | | |
|------|------|-----|--------|---|--------------------|
| C 24 | 33 | 4.1 | 395 | 4 | AA185726 Human pol |
| C 25 | 33 | 4.1 | 3971 | 7 | AA47845 Haemophil |
| C 26 | 33 | 4.1 | 4942 | 4 | AB105954 Drosophi |
| C 27 | 33 | 4.1 | 30121 | 2 | AA142063-18 |
| C 28 | 33 | 4.1 | 110000 | 2 | AA142063-17 |
| C 29 | 32.8 | 4.1 | 615 | 6 | ABQ26306 Oligonuc |
| C 30 | 32.8 | 4.1 | 615 | 6 | ABQ26307 Oligonuc |
| C 31 | 32.8 | 4.1 | 1109 | 3 | AAQ45218 Arabidops |
| C 32 | 32.8 | 4.1 | 1115 | 3 | AAQ35493 Arabidops |
| C 33 | 32.8 | 4.1 | 4580 | 5 | AAH24065 Yeast Abd |
| C 34 | 32.6 | 4.1 | 309 | 4 | AAH3611 S. epider |
| C 35 | 32.6 | 4.1 | 648 | 7 | ABT15000 Pathogen |
| C 36 | 32.6 | 4.1 | 651 | 4 | AAH3615 S. epider |
| C 37 | 32.6 | 4.1 | 825 | 6 | ABN92979 Staphyloc |
| C 38 | 32.6 | 4.1 | 3549 | 4 | AAH54658 S. epider |
| C 39 | 32.6 | 4.1 | 349980 | 5 | AAH66431 Pyrococu |
| C 40 | 32.4 | 4.1 | 1308 | 7 | ACA36267 Prokaryot |
| C 41 | 32.4 | 4.1 | 1494 | 4 | AAH52644 E. coli D |
| C 42 | 32.4 | 4.1 | 1494 | 7 | ACA32678 Prokaryot |
| C 43 | 32.4 | 4.1 | 1620 | 7 | AAQ03541 Chinese h |
| C 44 | 32.4 | 4.1 | 1830 | 7 | ACA54254 Prokaryot |
| C 45 | 32.4 | 4.1 | 2742 | 5 | AAH93830 DNA encod |

ALIGNMENTS

| | | |
|----------|---|------------------------|
| RESULT 1 | ACA53528 | standard; DNA; 717 BP. |
| XX | ACA53528: | |
| XX | 19-JUN-2003 (first entry) | |
| DE | Prokaryotic essential gene #35185. | |
| XX | Antisense; ds; prokaryotic essential gene; cell proliferation; | |
| XX | drug design; gene. | |
| XX | Vibrio cholerae. | |
| XX | WO200277183-A2. | |
| XX | 03-OCT-2002. | |
| XX | 21-MAR-2002; 2002WC-US009107. | |
| XX | 21-MAR-2001; 2001US-00815242. | |
| XX | 06-SEP-2001; 2001US-00948993. | |
| XX | 25-OCT-2001; 2001US-0342823P. | |
| XX | 08-FEB-2002; 2002US-00072851. | |
| XX | 06-MAR-2002; 2002US-0362699P. | |
| PA | (ELIT-) ELITRA PHARM INC. | |
| XX | Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zykkind UW; | |
| XX | Wall D, Trawick JD, Carr GV, Yamamoto R, Forsyth RA, Xu HH; | |
| XX | WPI; 2003-029926/02. | |
| XX | P-PDSB; ABV49658. | |
| PT | New antisense nucleic acids, useful for identifying proteins or screening | |
| PT | for homologous nucleic acids required for cellular proliferation to | |
| PT | isolate candidate molecules for rational drug discovery programs. | |
| XX | Claim 14; SEQ ID NO 41398; 1766bp; English. | |
| CC | The invention relates to an isolated nucleic acid comprising any one of | |
| CC | the 6213 antisense sequences given in the specification where expression | |
| CC | of the nucleic acid inhibits proliferation of a cell. Also included are: | |
| CC | (1) a vector comprising a promoter operably linked to the nucleic acid | |
| CC | encoding a polypeptide whose expression is inhibited by the antisense | |

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation; (7) identifying a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

Sequence 717 BP, 175 A, 165 C, 187 G, 190 T, 0 U, 0 Other;

Query Match 50.9%; Score 405.8; DB 7; Length 717;

Best Local Similarity 73.1%; Pred. No. 4e-119; Matches 521; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 85 ATGATTTGCTATCGATATGATTTGCTGCGCTTGGCCGCAATCTATCCAGGTTTAA 144
 DB 1 ATGATTTGCTATCGATATGATTTGCTGCGCTTGGCCGCAATCTATCCAGGTTTAA 60
 QY 145 CTGGTGAACGCTTATTCGAAATGGCCAGATTGTGTAATGGCCATTGCGCGCG 204
 DB 61 TTGGTCAACGCGCAATCCAAACCGCTGATGGCGCTTGGCGCAATCCCGTGTGT 120
 QY 205 ATAGTGTATGACGAAATGATGACCGCTCATGGGAGAACTGTGATGAGGATTTGAT 264
 DB 121 TGGGTGTATGATGAGATCTCAGCCGCAAGGTGGGAAACCGCTGACAAAGATTTGAT 180
 QY 265 GCAGCGAGACGATTTTGTGGCAAAAGTCCATCTTATCCATTTTATGACGCAAT 324
 DB 181 TCCGCGCGTGGGCGATCTGCGCGCAAAATCCATCTTATCCCAATCTTATGAGTAT 240
 QY 325 CCGCTGGTGTATGAGCAACCCCAACCGCATCCGAATGTTGAGTCAATTTTCCAT 384
 DB 241 CTTTGTGTGACGATATCTTAAACGCAATCCAAAGGTTGAGCATCTGATTTTCTAC 300
 QY 385 TACGCTTATTAACCGCTGATGATCAAAAGATGAGATTTTGTATGACCCGAG 444
 DB 301 TATGCCCTGTCTATCTTCAATCCAGTCAAGCATGAGATGAGATGATGAGATGAGC 360
 QY 445 CCGCTTATGTTGTTGATCTTCACTTACTTCAAAAGAAATATGCGCTGCTTTGAT 504
 DB 361 CGTGTATATGTTGTTTCCACTAGAGATGATTTGCAAGGTTTCAAGTTCTGCGTTTAT 420
 QY 505 CATGTGCGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 564
 DB 421 CATGTGCGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 565 GTGGTACTATTTTCAATTAAGAAAAAGTTTATGATGATGATGATGATGATGATGATGAT 624
 DB 481 GTGGTACTATTTTCAATTAAGAAAAAGTTTATGATGATGATGATGATGATGATGATGAT 540
 QY 625 AATTTTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684
 DB 541 AATTTTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 685 ATGCTAGTACCATTAATAATGCGGATCTGTGAAGCAAGCAACCAAGCCACC 744
 DB 601 ATGCTAGTACCATTAATAATGCGGATCTGTGAAGCAAGCAACCAAGCCACC 660
 QY 745 GTTATGCTTTTCCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 797
 DB 661 GTTATGCTTTTCCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 713

RESULT 2
 AAD57245/c
 ID AAD57245 standard; cDNA, 6065 BP.
 XX
 AC AAD57245;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human CGDD-25 cDNA.
 XX
 KW Human; cell growth, differentiation and death protein; CGDD; leukaemia;
 KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
 KW muscular disorder; myotonic dystrophy; catatonias; endocrine disorder;
 KW diabetes; Graves' disease; cancer; immunological disorder; scleroderma;
 KW systemic lupus erythematosus; allergy; Crohn's disease; renal disorder;
 KW gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis;
 KW cardiovascular disorder; atherosclerosis; hepatic disease; transgenic;
 KW transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic;
 KW dermatological; immunosuppressive; cerebroprotective; anticonvulsant;
 KW antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiac;
 KW protozoicide; nootropic; gene; ss.
 KW
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1442..4129
 FT /*tag= a
 FT /product= "Human CGDD-25 protein"
 EN MO203050253-A2.
 XX
 PD 19-JUN-2003.
 XX
 PE 04-DEC-2002; 2002WC-US039133.
 XX
 PR 07-DEC-2001; 2001US-0340747P.
 PR 20-DEC-2001; 2001US-0342761P.
 PR 15-JAN-2002; 2002US-0349705P.
 PR 06-FEB-2002; 2002US-0354764P.
 PR 12-FEB-2002; 2002US-0356216P.
 XX
 RA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Griffin JA, Rankumar J, Emerling BM, Kable AE, Elliott VS,
 PI Margus JP, Baughn MR, Gorvay AE, Yue H, Lee EA, Becha SD, Tang YT,
 PI Tran UK, Swarnakar A, Lee S, Ison CH, Hafalla AJA, Tran B,
 PI Sprague WM, Lee SY, Khare R, Gandhi AR, Gietzen KJ, Bhatia U,
 PI Burdill JD, Blake UT, Ho A, Zheng W,
 XX
 DR WPI; 2003-532903/50.
 DR P-PSDB; AAE37936.
 XX
 PT New CGDD polypeptides, useful for diagnosing, preventing, and treating
 PT disorders associated with an abnormal expression or activity of CGDD,
 PT e.g. neuromuscular, immunological, cardiovascular disorders, cancer
 PT and/or infections.
 XX
 PS Claim 5; Page 289-290; 299pp; English.
 CC The present invention relates to novel cell growth, differentiation and
 CC death (CGDD) proteins and polynucleotides encoding them. The sequences of
 CC the invention are useful in diagnosing, preventing and treating disorders
 CC associated with an abnormal expression or activity of CGDD such as
 CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's

CC disease), muscular disorders (e.g. myotonic dystrophy, catatonias),
 CC endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.
 CC leukemia, cervical or breast cancers), immunological disorders (e.g.
 CC scleroderma, systemic lupus erythematosus, allergies), gastrointestinal
 CC disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's
 CC syndrome), infections (e.g. viral, bacterial, fungal, parasitic,
 CC protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis)
 CC and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to
 CC create humanised animals or transgenic animals to model human diseases.
 CC The invention is also used in gene therapy. The present sequence is human
 CC CGDD-25 cDNA

SO Sequence 6065 BP, 2249 A, 1216 C, 1149 G, 1451 T, 0 U, 0 Other;

Query Match 4.8%; Score 38.2; DB 8; Length 6065;

Best Local Similarity 52.9%; Pred. No. 0.62; Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 240 AGAAGCTGTGATGAGATTTTGATGACGACGACGATTTTGTGGCAAAAGTCCA 299
 DB 290 AGGACCTTTCATGCGCATTTTAAAGCTGTGAGATTTTGTGAGAAAGTCT 231
 QY 300 TACTTATCTAATTTATGACGATCGCTGTGTATGSCAACCCCAACGCGATCGAA 359
 DB 230 GCGTTTTCAGTCAATCAGTCTCCCGCTTTTGTGTGCTTAAATGATTTGAC 171
 QY 360 TGGTTGAGTGTGATGATTTCCCATTCAGCTTTAT 394
 DB 170 TTTTTCAGAGACTTTCATTTCTGTACATGAT 136

RESULT 3

ADA71938/c
 ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;
 XX gene; de.

XX Oryza sativa.

XX WO2003000838-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-1B001105.

XX 22-JUN-2001; 2001WO-1B001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 XX Katsir I, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to
 XX pathogenic infection for conferring resistance or tolerance to a plant to
 XX bacterial, fungal or viral infection by determining or detecting plant
 XX gene expression.

XX Claim 27; SEQ ID NO 5263; 899bp; English.

XX The present invention relates to a method (M1) for identifying genes
 XX involved in plant resistance or response to pathogenic infection. M1
 XX comprises identifying a gene whose expression is significantly altered in
 XX the incompatible interaction of plant gene expression relative to
 XX expression of the gene in an uninfected plant, in a mutant plant that

CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.

SO Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.8%; Score 38; DB 7; Length 2000;

Best Local Similarity 9.6%; Pred. No. 0.42; Matches 74; Conservative 335; Mismatches 355; Indels 4; Gaps 2;

QY 19 TATTACATTTGACCATCATGACATCTTACCTTAATAAGCCGTTTATTAGGAA 78
 DB 855 TMMWRMTTTCYCAATCAKCKYKMTTMAAGMTATSRMRWMAAGMRKRYMKR 796
 QY 79 GCGATTATGATGTGACATGATGATGATGATGATGATGATGATGATGATGATGAT 137
 DB 795 AYMMWRMRCKAGARMKRMYRMKRYMYRYMYRYMYRYMYRYMYRYMYRYMYRY 736
 QY 138 GATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 197
 DB 735 MRMWMTTTCYCAATCAKCKYKMTTMAAGMTATSRMRWMAAGMRKRYMKR 676
 QY 198 TGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 257
 DB 675 ARMSKRRKMAAGASMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSK 616
 QY 258 TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 317
 DB 615 YKMSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMT 556
 QY 318 CAGCGATCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 374
 DB 555 WKCTWRRCMYRMYRYMYRYMYRYMYRYMYRYMYRYMYRYMYRYMYRYMYRYMY 496
 QY 375 TATTCCATTAAGCTTATTAACCCGCGGATGATGATGATGATGATGATGATGAT 434
 DB 495 CSTMMRYGYCKACKCCYACWMAAASGMMYRYRYRYRYRYRYRYRYRYRYRY 436
 QY 435 CGAAGCGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 494
 DB 435 YGAKGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 376
 QY 495 GCGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 554
 DB 375 WYTYAATYMYRY 316
 QY 555 ATACACATCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 614
 DB 315 CWRYATCYGCCYRKGWYSRSMRTAGMKRMSRMSRMSRMSRMSRMSRMSRMS 256
 QY 615 TGCTTACGCAAAATTTGGCGTGAAGTATGATGATGATGATGATGATGATGAT 674
 DB 255 YGMASSGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196
 QY 675 TACCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 734
 DB 195 KMSKSGSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMTSMT 136
 QY 735 ACCAGCACCGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 782
 DB 135 KYSCGCTKYSGYRYCYKMYKYKYKYKYKYKYKYKYKYKYKYKYKYKYKYKY 88

RESULT 4

ADA71938
 ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX 20-NOV-2003 (first entry)

PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234979P.
 PR 25-SEP-2000; 2000US-0234980P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235634P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239355P.
 PR 13-OCT-2000; 2000US-023937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0251160P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251866P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254907P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM,
 DR WPI; 2001-541565/60.
 XX
 XX
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides.
 PT useful for preventing, diagnosing and/or treating nervous system cancers
 PT and metastases.
 XX
 XX
 PS Disclosure; SEQ ID NO 11193; 1701pp + Sequence Listing; English.
 XX
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA1678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Notes: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences
 XX
 SQ Sequence 438 BP; 130 A; 78 C; 104 G; 126 T; 0 U; 0 Other;
 Query Match 4.6%; Score 36.8; DB 5; Length 438;
 Best Local Similarity 59.6%; Pred. No. 0.48; Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
 Db 357 GAATGGTTGAGTGTGATGATTTCCATTACGCTTTATTAAACCGTGAAATGTCAAQA 416
 204 GGAAGATGAGTGTGATGATTTCCATTACGCTTTATTAAACCGTGAAATGTCAAQA 263
 QY 417 AATGAGATTTTGTATGATGATTTCCATTACGCTTTATTAAACCGTGAAATGTCAAQA 460
 264 AACCTAACATTTTGTATGATTTCTGGGTTTGTCTTTTAACTTTG 307
 Db
 RESULT 6
 ABA1863
 ID ABA1863 standard; DNA; 438 BP.
 AC ABA1863;
 XX
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Human nervous system related polynucleotide SEQ ID NO 11194.
 XX
 XX Human; noctropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnary;
 KW antiparkinsonian; antistoking; antianaemic; antiarthritic; cancer;

KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antidiuretic; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.

XX WO200159063-A2.

PD 16-AUG-2001.

PF 17-JAN-2001; 2001WO-US001334.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205151P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214866P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 11-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0217496P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224418P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225131P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225265P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234397P.

PR 25-SEP-2000; 2000US-0234988P.

PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236882P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0237049P.

PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.

PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.

PR 08-NOV-2000; 2000US-0246525P.

PR 08-NOV-2000; 2000US-0246526P.

PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.

PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.

PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246612P.

PR 17-NOV-2000; 2000US-0249212P.

PR 17-NOV-2000; 2000US-0249213P.

PR 17-NOV-2000; 2000US-0249214P.

PR 17-NOV-2000; 2000US-0249215P.

PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.

PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249246P.

PR 17-NOV-2000; 2000US-0249265P.

PR 17-NOV-2000; 2000US-0249297P.

PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250391P.

PR 01-DEC-2000; 2000US-0251160P.

PR 05-DEC-2000; 2000US-0251030P.

PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.

PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251899P.

PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 11194; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pdb/published_pdb_sequences
XX
SQ Sequence 438 BP; 130 A; 78 C; 104 G; 126 T; 0 U; 0 Other;
Query Match 4.6%; Score 36.8; DB 5; Length 438;
Best Local Similarity 59.6%; Pred. No. 0.48;
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 357 GATGGTTCAGAGTCAGTATTTCCCATTCAGCCTTATTAACCCGGAATGTCAACA 416
DB 204 GGAAGATATGAGTCGAGTATTCACATTAACGCTTATTTCAACTGGAAATGATAACA 263
QY 417 AATGAAGATTTTGGTATGACCCGAGCGCGCTAATGGTTG 460
DB 264 AACCTAACATTTGGTAGTCTCGGTTTGTCTTTTAACTTG 307
RESULT 7
ABA18866
ID ABA18866 standard; DNA; 439 BP.
XX
XX ABA18866;
AC
XX
XX 23-JAN-2002 (first entry)
DT
XX
XX Human nervous system related polynucleotide SEQ ID NO 11197.
DE
XX Human; noctropic; neuroprotective; cytoprotic; dermatological; virologic;
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnary;
XX antiparkinsonian; antischizoid; antianemic; antiautistic; cancer;
XX antineuritic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200159063-A2.
PN
XX
XX 16-AUG-2001.
PD
XX

PF 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218280P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232081P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234232P.
PR 21-SEP-2000; 2000US-0234233P.
PR 25-SEP-2000; 2000US-0234574P.
PR 25-SEP-2000; 2000US-0234575P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.

| | | |
|----|---|-----------------------|
| PR | 29-SEP-2000; | 2000US-0236370P. |
| PR | 02-OCT-2000; | 2000US-0236802P. |
| PR | 02-OCT-2000; | 2000US-0237037P. |
| PR | 02-OCT-2000; | 2000US-0237038P. |
| PR | 02-OCT-2000; | 2000US-0237039P. |
| PR | 13-OCT-2000; | 2000US-0237040P. |
| PR | 13-OCT-2000; | 2000US-0239935P. |
| PR | 13-OCT-2000; | 2000US-0239937P. |
| PR | 20-OCT-2000; | 2000US-0240960P. |
| PR | 20-OCT-2000; | 2000US-0241785P. |
| PR | 20-OCT-2000; | 2000US-0241786P. |
| PR | 20-OCT-2000; | 2000US-0241787P. |
| PR | 20-OCT-2000; | 2000US-0241808P. |
| PR | 20-OCT-2000; | 2000US-0241809P. |
| PR | 20-OCT-2000; | 2000US-0241826P. |
| PR | 01-NOV-2000; | 2000US-0244617P. |
| PR | 08-NOV-2000; | 2000US-0246474P. |
| PR | 08-NOV-2000; | 2000US-0246475P. |
| PR | 08-NOV-2000; | 2000US-0246476P. |
| PR | 08-NOV-2000; | 2000US-0246477P. |
| PR | 08-NOV-2000; | 2000US-0246478P. |
| PR | 08-NOV-2000; | 2000US-0246523P. |
| PR | 08-NOV-2000; | 2000US-0246524P. |
| PR | 08-NOV-2000; | 2000US-0246525P. |
| PR | 08-NOV-2000; | 2000US-0246526P. |
| PR | 08-NOV-2000; | 2000US-0246527P. |
| PR | 08-NOV-2000; | 2000US-0246528P. |
| PR | 08-NOV-2000; | 2000US-0246532P. |
| PR | 08-NOV-2000; | 2000US-0246609P. |
| PR | 08-NOV-2000; | 2000US-0246610P. |
| PR | 08-NOV-2000; | 2000US-0246611P. |
| PR | 08-NOV-2000; | 2000US-0246613P. |
| PR | 17-NOV-2000; | 2000US-0249207P. |
| PR | 17-NOV-2000; | 2000US-0249208P. |
| PR | 17-NOV-2000; | 2000US-0249209P. |
| PR | 17-NOV-2000; | 2000US-0249210P. |
| PR | 17-NOV-2000; | 2000US-0249211P. |
| PR | 17-NOV-2000; | 2000US-0249212P. |
| PR | 17-NOV-2000; | 2000US-0249213P. |
| PR | 17-NOV-2000; | 2000US-0249214P. |
| PR | 17-NOV-2000; | 2000US-0249215P. |
| PR | 17-NOV-2000; | 2000US-0249216P. |
| PR | 17-NOV-2000; | 2000US-0249217P. |
| PR | 17-NOV-2000; | 2000US-0249218P. |
| PR | 17-NOV-2000; | 2000US-0249244P. |
| PR | 17-NOV-2000; | 2000US-0249245P. |
| PR | 17-NOV-2000; | 2000US-0249264P. |
| PR | 17-NOV-2000; | 2000US-0249265P. |
| PR | 17-NOV-2000; | 2000US-0249297P. |
| PR | 17-NOV-2000; | 2000US-0249299P. |
| PR | 17-NOV-2000; | 2000US-0249300P. |
| PR | 01-DEC-2000; | 2000US-0250391P. |
| PR | 01-DEC-2000; | 2000US-0251160P. |
| PR | 05-DEC-2000; | 2000US-0251030P. |
| PR | 05-DEC-2000; | 2000US-0251988P. |
| PR | 05-DEC-2000; | 2000US-0256719P. |
| PR | 06-DEC-2000; | 2000US-0251479P. |
| PR | 08-DEC-2000; | 2000US-0251856P. |
| PR | 08-DEC-2000; | 2000US-0251868P. |
| PR | 08-DEC-2000; | 2000US-0251869P. |
| PR | 08-DEC-2000; | 2000US-0251989P. |
| PR | 08-DEC-2000; | 2000US-0251990P. |
| PR | 11-DEC-2000; | 2000US-0254097P. |
| PR | 05-JAN-2001; | 2001US-025678P. |
| XX | | |
| PA | (HUMA-) | HUMAN GENOME SCI INC. |
| XX | | |
| PI | Rosen CA, | Barash SC, Ruben SM; |
| XX | | |
| DR | WPI, | 2001-541565/60. |
| XX | | |
| NT | Nucleic acids encoding 3324 human nervous system antigen polypeptides | |

| PT | useful for preventing, diagnosing and/or treating nervous system cancers |
|------------|---|
| PI | and metacases. |
| PS | Disclosure; SEQ ID NO 11197; 1701bp + Sequence Listing; English. |
| XX | The invention relates to novel genes (AB011004-AB021534) and proteins |
| CC | (AB011678-AB0118001) useful for preventing, treating or ameliorating |
| CC | medical conditions e.g. by protein or gene therapy. The genes are |
| CC | isolated from a range of human tissues disclosed in the specification. |
| CC | The nucleic acids, proteins, antibodies and (ant)agonists are useful in |
| CC | the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and |
| CC | ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, |
| CC | breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune |
| CC | disorders e.g. Addison's disease, allergies, autoimmune haemolytic |
| CC | anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, |
| CC | multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) |
| CC | cardiovascular disorders such as myocardial ischaemias; (d) wound healing |
| CC | infectious diseases such as viral, bacterial, fungal and parasitic |
| CC | infections. Note: The sequence data for this patent did not form part of |
| CC | the printed specification, but was obtained in electronic format directly |
| CC | from WIPO at ftp.wipo.int/pub/published_pct_sequences |
| XX | |
| SQ | Sequence 439 BP; 130 A; 79 C; 104 G; 126 T; 0 U; 0 Other; |
| | Query Match 4.6%; Score 36.8; DB 5; Length 439; |
| | Best Local Similarity 59.6%; Pred. No. 0.48; Mismatches 0; Gaps 0; |
| | Matches 62; Conservative 0; Indels 42; Indels 0; Gaps 0; |
| OY | 357 GAATGTTGAGAGTCTCAGTATTTCCATTACGCTTATTTAAACCGTGAAATGCAACA 416 |
| DB | 205 GGAAGAGTAGAGTCTCAGATTCACATTAACGCTTATTTCAACTTGGAATGATTAACA 264 |
| OY | 417 AATGAGATTTTGGTATCGACCCCGAGCGCGCTAATGGTTG 460 |
| DB | 265 AACCTAACATTTTGGTAGTCTCTGGGATTGCTCTTTAACTTIG 308 |
| RESULT 8 | |
| AB088161/c | |
| ID | AB088161 standard; cDNA; 73465 BP. |
| XX | AB088161; |
| AC | |
| XX | 18-SEP-2002 (first entry) |
| DT | |
| XX | Human osteoblast differentiation related cDNA SEQ ID NO 68. |
| DE | |
| XX | Human; osteoblast; stem cell differentiation; bone tissue deposition; |
| KW | osteoporosis; osteopathic; ss. |
| KW | |
| OS | Homo sapiens. |
| OS | |
| XX | W0200250301-A2. |
| PN | |
| XX | 27-JUN-2002. |
| PD | |
| XX | 18-DEC-2001; 2001WO-US048276. |
| PF | |
| PR | 18-DEC-2000; 2000US-0255882P. |
| PR | 24-APR-2001; 2001US-0285691P. |
| XX | |
| PA | (GENE-) GENE LOGIC INC. |
| PA | (PROC) PROCTER & GAMBLE CO. |
| XX | |
| PI | Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A; |
| PI | Mertz L; |
| XX | WPI; 2002-557663/59. |
| DR | |
| XX | |
| PT | Use of genes and their expression profiles associated with osteoblast |
| PT | differentiation for screening modulators bone formation; for diagnosing |
| PT | or treating e.g. osteoporosis, or as markers for the differentiation |

PT process.
XX
PS Claim 1; SEQ ID NO 68; 789p + Sequence Listing; English.
XX
CC The invention relates to genes and their expression profiles are used
CC for: (a) screening modulators of precursor stem cell differentiation into
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
CC deposition of bone tissue, abnormal rate of osteoblast formation or
CC osteoporosis; or (c) treating or monitoring treatment of the conditions
CC cited in (b), or monitoring the progression of bone tissue deposition.
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
CC induced abnormalities in bone formation or bone loss, conditions that
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
CC or fibrous dysplasia. The present sequence is that of an osteoblast
CC differentiation associated cDNA marker of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/publ/published_pct_sequences
XX
SQ Sequence 73465 BP; 21307 A; 15130 C; 15840 G; 21188 T; 0 U; 0 Other;
Query Match 4.5%; Score 36; DB 6; Length 73465;
Best Local Similarity 51.2%; Pred. No. 11;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 20 ATTACATTGACATCATCATGCACTTACTTAATAAGCCCGTTTATTAGGGAAG 79
DB 72386 ATACATATTTTACATTAATTAATGACAAACGAGATTAATGAGTTCTTGTGAT 72327
QY 80 CCATTATGATTTGTCACTATCATGATGATTTGTGCGCTTGGCGGAATCATTCAGG 139
DB 72326 GCAATTAAGAGAGTGTCTTCAATAGATTAATTAATTTGTATTTCTTATCTTACTGC 72267
QY 140 TTTTACTGGTGAACGCTCTAATCCAAATCGCGCAGATTGTGT 183
DB 72266 TTGCTGCGCTGTAAACGTTTATAAAGAGTTCCCTTTGGGGT 72223
RESULT 9
ADC90659
ID ADC90659 strand; DNA; 552 BP.
XX
AC ADC90659;
XX
DT 01-JAN-2004 (first entry)
XX
DE E. faecium DNA sequence SEQ ID 286.
XX
KW ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
XX abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
XX 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI: 2003-799836/75.
XX P-PSDB; ADC94313.
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and

PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX
PS Example 1; SEQ ID NO 286; 243bp; English.
XX
CC The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridizing to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids are useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium nucleic acids.
XX
SQ Sequence 552 BP; 174 A; 96 C; 135 G; 147 T; 0 U; 0 Other;
Query Match 4.5%; Score 35.6; DB 9; Length 552;
Best Local Similarity 51.2%; Pred. No. 1.3;
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 237 TGGAGAACCTGTGATGAGGATTTTATGACAGGAGACGATTTTGTGCGCAAAAGT 296
DB 294 TAGTCAACATCTCAGAAATATCTATGTTTGGGTGGAGAGCGGATTTTTCMAAGACT 353
QY 297 CCATTCTTATCTTAATTTATTCAGCGATCGGCTGTTATGCGAACCCCAAGCCATCC 356
DB 354 TCTTCCCTTACCGCGTGCAGATCTGGCGGACATGATGATGACGCTTGAAGAGATAC 413
QY 357 GAATGGTTGGAGTGTGATTTTCCATTACGCTTTATTAA 398
DB 414 GTATATGGGGATATCATTTTTCAGACCTTGGCTTATAGA 455
RESULT 10
ADB76952
ID ADB76952 strand; DNA; 1119 BP.
XX
AC ADB76952;
XX
DT 04-DEC-2003 (first entry)
XX
DE Mouse CLCA4 gene SEQ ID NO:20.
XX
KW respiratory; gastrointestinal; nephrotropic; respiratory disease;
XX nephritis; digestive disease; mouse; CLCA4; ds; gene.
XX
OS Mus sp.
XX
PN Key Location/Qualifiers
XX CDS 1..1119
XX FT /product= "CLCA4"
XX FT /note= "No start/stop codon given"
XX
XX MO2003062426-A1.
XX
XX 31-JUL-2003.
XX
XX 20-JAN-2003; 2003MO-JP000408.
XX
XX 21-JAN-2002; 2002JP-00012180.
XX 21-FEB-2002; 2002JP-0004981.
XX 07-MAR-2002; 2002JP-00061668.

```

XX (TAKE ) TAKEDA CHEM IND LTD.
PA
XX Nakanishi A, Morita S;
PI
XX WPI; 2003-598752/56.
DR
XX P-PSDB; ADB36951.
DR
XX Proteins for treatment, diagnosis and prevention of respiratory disease,
PT nephritis, digestive diseases.
XX
XX PS Claim 12; Page 104; 113pp; Japanese.
XX
CC The invention relates to a novel protein comprising a sequence selected
CC from two protein sequences each 924 amino acids in length. A protein of
CC the invention has respiratory, gastrointestinal, and nephrotropic
CC activity. The protein is useful for treatment, diagnosis and prevention
CC of respiratory disease, nephritis, and digestive diseases. The present
CC sequence encodes mouse ClCA4.
XX
SQ Sequence 1119 BP; 352 A; 245 C; 261 G; 261 T; 0 U; 0 Other;
Query Match 4.4%; Score 35.4; DB 9; Length 1119;
Best Local Similarity 50.3%; Pred. No. 2.1;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 601 GCGGATTTATTTGATGCTTACGCCAATTGGCGTGAAGTTATGCAATGACATTAA 660
DB 898 GCTGATGCTCACTGAGAGTCCCTTCAGCTTGAAGAGAGGTGCTGTGACCAACAT 957
QY 661 CGCGGCTTGAATCAATACCGGGGTGATGTCAGTACCAATTAATGGCCGATCTGTAA 720
DB 958 CGCTGGCTGACGACACTGCTGTTATTGACAGCACCCTGGAGAAAGACACATCTTCT 1017
QY 721 GCGAAGAGAGCCAAACCGACCGCTTATGCTTCCAGTCATGAAGTCAC 773
DB 1018 GTACACCTGAGCAACACAGCCCTGCATTCATCTCAGGATCCCAAGAAAC 1070

RESULT 11
AAQ49902/C
ID AAQ49902 standard; cDNA; 4368 BP.
XX
XX AAQ49902;
AC
XX 14-MAY-2003 (revised)
DT 04-MAY-1994 (first entry)
XX
DE Glutamic acid receptor.
XX
XX Glutamic acid receptor; nerve; synapse; synapse plasticity; cerebellum;
KW cell necrosis; ischemia; ss.
XX
OS Mus musculus.
XX
XX Key Location/Qualifiers
FT CDS 1..4368
FT FT /*tag= a
FT FT /product= "glutamic_acid_receptor"
XX
XX JP05239098-A.
XX
XX 17-SEP-1993.
XX
XX 26-FEB-1992; 92JP-00039563.
XX
XX 26-FEB-1992; 92JP-00039563.
XX
XX (MITU ) MITSUBISHI KASEI CORP.
XX
XX WPI; 1993-331429/42.
XX
XX P-PSDB; AAR42054.
XX

```

```

PT New protein for elucidation of nerve information transfer by synapse -
PT comprises glutamic acid receptor of specified aminoacid sequence.
XX
XX Claim 1; Page 11; 15pp; Japanese.
XX
CC The sequence was transformed into E. coli several clones coding the new
CC glutamic acid receptor were obtained. The sequence is useful for the
CC elucidation of nerve information transfer by synapse, expression of
CC synapse plasticity and nerve cell necrosis caused by cerebral ischemia
CC and also for the development of new drugs. In the specification the
CC length of the sequence is given as 4392 bp The actual number of bases is
CC 4368. The sequence may have a line missing or has been numbered
CC incorrectly. (Updated on 14-MAY-2003 to correct PS field.)
XX
SQ Sequence 4368 BP; 1151 A; 1127 C; 1090 G; 1000 T; 0 U; 0 Other;
Query Match 4.4%; Score 35.2; DB 2; Length 4368;
Best Local Similarity 49.0%; Pred. No. 4.8;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 234 TGTGAGAGACCTGTGCATGAGATTTGATTCAGCAGCAGACGATTTTGTGGCAAAA 293
DB 553 TGATGAAGTCTCTGTACCAAGGAGATGTTGTGACCAAGAGACATGCCAGTCAT 494
QY 294 AGTCACATCTTACCTAATTTTATCAGCAGATCCGCTGTTGATGCAACCCCAACCGCA 353
DB 493 AGTCTGATATCTTTCAGCATGACCGTGTCTGCTGATGGAAGCTCCAACTGGA 434
QY 354 TCCGATGTGTGAGTGTGATGATTTCCATTCAGCTTTATTAACCCGTGAATGTCAA 413
DB 433 AGAAGCGGAGTGTGATGATCCCTTGCACGCATGATCATGATGACACCCCATGATGCCA 374
QY 414 ACAATAGAGAA 425
DB 373 AGATGGGATGA 362

RESULT 12
AB199832/C
ID AB199832 standard; cDNA; 4512 BP.
XX
XX AB199832;
AC
XX 07-MAR-2002 (first entry)
DT
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:952.
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX Mus musculus.
XX
XX WO200188188-A2.
XX
XX 22-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-JP004192.
XX
XX 18-MAY-2000; 2000JP-00145977.
XX
XX (UNIV-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
PI
XX WPI; 2002-034733/04.
XX
XX P-PSDB; ABB57341.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX expression levels of particular genes defined in the specification or by
XX determining the expression profile of a gene group comprising these
XX genes.
XX
XX Claim 2; Page 2406-2415; 2690pp; English.
XX

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:27 ; Search time 2265.3 Seconds
(without alignments)
10519.601 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798
Sequence: 1 atgttgcacacaaagtctta.....ttcaacctgttacagagc 798

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 41.6 | 5.2 | 1048 | 9 | AL545951 AL545951 |
| 2 | 41 | 5.1 | 486 | 14 | CF088325 QM18203 |
| 3 | 40.8 | 5.1 | 746 | 29 | CC876255 ZMMBBD019 |
| 4 | 40.8 | 5.1 | 826 | 28 | BZ723869 PUBNSO9TD |

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 5 | 40.8 | 5.1 | 919 | 28 | CC389867 PUB1W66MD |
| 6 | 40.4 | 5.1 | 438 | 14 | CF090249 QM3P15.Y |
| 7 | 40.4 | 5.1 | 464 | 14 | CF087071 QM12N04 |
| 8 | 40.2 | 5.0 | 638 | 28 | AZ396257 IM0160C17 |
| 9 | 39.8 | 5.0 | 401 | 9 | AJ533921 AJ533921 |
| 10 | 39.8 | 5.0 | 435 | 14 | CF090254 QM3P22.Y |
| 11 | 39.6 | 5.0 | 437 | 13 | BX500256 DKF2P779J |
| 12 | 39.4 | 4.9 | 447 | 14 | CF091479 QM6E22.Y |
| 13 | 39.4 | 4.9 | 454 | 14 | CF090121 QM3J13.Y |
| 14 | 39 | 4.9 | 401 | 14 | CF096847 QM2P11.Y |
| 15 | 39 | 4.9 | 451 | 9 | AJ540178 AJ540178 |
| 16 | 38.8 | 4.9 | 418 | 14 | CF090542 QM6D12.Y |
| 17 | 38.8 | 4.9 | 427 | 14 | CF091120 QM7C23.Y |
| 18 | 38.8 | 4.9 | 437 | 14 | CF090736 QM5P20.Y |
| 19 | 38.8 | 4.9 | 440 | 14 | CF088879 QM20E19 |
| 20 | 38.8 | 4.9 | 452 | 14 | CF087459 QM15B24 |
| 21 | 38.8 | 4.9 | 454 | 14 | CF090001 QM3E04.Y |
| 22 | 38.8 | 4.9 | 468 | 14 | CF091500 QM6G22.Y |
| 23 | 38.8 | 4.9 | 472 | 14 | CF088587 QM1E02.Y |
| 24 | 38.8 | 4.9 | 487 | 14 | CF087094 QM12014 |
| 25 | 38.8 | 4.9 | 520 | 14 | CF090018 QM3B21.Y |
| 26 | 38.6 | 4.8 | 274 | 14 | CF096559 QM23J07.Y |
| 27 | 38.2 | 4.8 | 417 | 14 | CF091407 QM6C12.Y |
| 28 | 38.2 | 4.8 | 429 | 14 | CF088180 QM1E05.Y |
| 29 | 38.2 | 4.8 | 433 | 14 | CF087501 QM15G17 |
| 30 | 38.2 | 4.8 | 432 | 14 | CF087987 QM17107 |
| 31 | 38.2 | 4.8 | 542 | 14 | CF089042 QM20C09 |
| 32 | 38.2 | 4.8 | 581 | 14 | CF091617 QM8W17.Y |
| 33 | 38.2 | 4.8 | 711 | 10 | BF693394 602081026 |
| 34 | 38.2 | 4.8 | 9744 | 29 | AY405009 Homo bab1 |
| 35 | 38 | 4.8 | 635 | 14 | CF094197 QM17C08 |
| 36 | 37.8 | 4.7 | 374 | 14 | CF087834 QM16P22 |
| 37 | 37.8 | 4.7 | 409 | 14 | CF090830 QM6E07.Y |
| 38 | 37.8 | 4.7 | 486 | 14 | CF090109 QM3I24.Y |
| 39 | 37.4 | 4.7 | 283 | 14 | CF095069 QM19P07 |
| 40 | 37.2 | 4.7 | 434 | 14 | CF089999 QM3B02.Y |
| 41 | 37.2 | 4.7 | 448 | 14 | CF087103 QM12P01 |
| 42 | 37.2 | 4.7 | 454 | 14 | CF088135 QM1E1A19 |
| 43 | 37.2 | 4.7 | 460 | 13 | BX485370 DKF2P6860 |
| 44 | 36.8 | 4.6 | 440 | 10 | BE081047 QV1-BT063 |
| 45 | 36.6 | 4.6 | 331 | 14 | CF088421 QM19B15 |

ALIGNMENTS

RESULT 1
AL545951 1048 bp mRNA linear EST 31-MAY-2003
DEFINITION AL545951 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
LOCUS CSOD10237B12 5-PRIME, mRNA sequence.
ACCESSION AL545951 GI:31267786
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Buthyryota; Euteleostomi; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1. (bases 1 to 1048)
Li, W.B., Gruber, C., Jessee, D., and Polya, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
On Feb 15, 2001 this sequence version replaced GI:12878614.
COMMENT
Contact: Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/InvitrogenCorporation 1600
Faraday Avenue Genoscope sequence ID: CSOD10237B06P1.

FEATURES
source

1..1048

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS01023YB12"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-collg (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 5.2%; Score 41.6; DB 9; Length 1048;
Best Local Similarity 47.3%; Pred. No. 0.5; 77; Indels 0; Gaps 0;
Matches 80; Conservative 12; Mismatches 77; Indels 0; Gaps 0;

QY 341 ACCCCAAACGCGATCCGATGTTGAGTGCAGTATTCCTTACGCTTATTAAACC 400
DB 536 AACCTCAAGAAACATTAAGCTTGAGTGCCTTAATTTTAAACGATTCATTAAC 595
QY 401 CGTGCATGCAACAAATGAGATTTTGTATCGACCCGCGCGCTAATGTTG 460
DB 596 GGTACTACCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 655
QY 461 ATCTTCATCTTACTCAAGAAAGAAATGCGCTGCTTGTATCATGT 509
DB 656 TTTTMTATTTATTTTAAATTAATTAATTAATTAATTAATTAATTTTAAATTT 704

RESULT 2 486 bp mRNA linear EST 22-JUL-2003
CF088325
LOCUS OHM18L03.YF.ab1 OH M sunflower H. argophyllus Helianthus argophyllus
DEFINITION OHM18L03.YF.ab1 OH M sunflower H. argophyllus Helianthus argophyllus
CDNA clone OHM18L03, mRNA sequence.

ACCESSION CF088325.1 GI:33127392
VERSION CF088325.1
KEYWORDS EST
SOURCE Helianthus argophyllus
ORGANISM Helianthus argophyllus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Asteroideae;
Heliantheae; Helianthus.

REFERENCE 1 (bases 1 to 486)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., Van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabough, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Letture and Sunflower ESTs from the Composite Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)

JOURNAL COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Aremundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OH_CA_Contig2435_2, see http://cgpdb.ucdavis.edu/
for details.

FEATURES
SOURCE Plate: OHM18 row: L column: 03.
Location/Qualifiers
1..486

/organism="Helianthus argophyllus"
/mol_type="mRNA"
/db_xref="taxon:73275"
/clone="OHM18L03"
/lab_host="E. coli"
/clone_lib="OH M sunflower H. argophyllus"
/note="Vector: pRCRNASFIAB; The library was constructed
from three different sources (seedling, root and leaf) of
RNA from a single genotype. cDNAs were pooled and
directionally cloned into a custom medium-copy vector.

Details of library construction can be obtained at
http://cgpdb.ucdavis.edu/"

Query Match 5.1%; Score 41; DB 14; Length 486;
Best Local Similarity 47.8%; Pred. No. 0.53;
Matches 119; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 460 GATCTTCATCTTACTTACCAAGAAAGAAATGCGCTGCTTTGATCATGTCGGCAATT 519
DB 62 GATACACATCTTAATTCACAAAGAGATGACCTCATATGTGACAAATCCAGT 121
QY 520 CAGCATGCGTGGCAAAATTAACGCGCTGCTGTAATACATCCGTGACTATTTC 579
DB 122 AAACTTCGTGGCAGAGGCTAGTCCGGGAAGAGAGATTGTGCTGAACATATTGAA 181
QY 580 TTAGAAAAGAGTTTATAGTGGCGGATATTTATGATCCCTACCGCAATTGGCGTGA 639
DB 182 AAGAAAACGGGAAGTTAATGCGCAACTATGCAATGAACACGATCCCGAGATA 241
QY 640 GTTATCGCATGACCATTAACGCGCTGATCATATCCGGGTGATCGTACATCAAT 699
DB 242 TCACTTTCGATTAAGTCTTGTGTGCTGACAAAGAGGATTCGTGAGTACCCCT 301
QY 700 AAAATGGCC 708
DB 302 ACCGTGGC 310

RESULT 3 746 bp DNA linear GSS 29-JUL-2003
CC876255
LOCUS ZMWBBD0194M17.r ZMWBBD Zea mays subsp. mays genomic clone
DEFINITION ZMWBBD0194M17.r, genomic survey sequence.
ACCESSION CC876255
VERSION CC876255
KEYWORDS GSS
SOURCE Zea mays subsp. mays (maize)
ORGANISM Zea mays subsp. mays (maize)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 746)
Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.,
and Wing, R.
Sequencing of the maize genome
Unpublished (2003)

JOURNAL COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

FEATURES
SOURCE PCR primers
FORWARD: 77
BACKWARD: M13r
Plate: 0194 row: M column: 17
Seq primer: M13r
Class: BAC ends.
Location/Qualifiers
1..746

/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMWBBD0194M17"
/lab_host="DH10B"
/clone_lib="ZMWBBD"
/note="Vector: pBel0AC11, Site_1: HindIII, Site_2:
HindIII; Zea mays L. ssp. mays"

ORIGIN

Query Match 5.1%; Score 40.8; DB 29; Length 746;
 Best Local Similarity 59.5%; Pred. No. 0.74;
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 388 GCTTATTAACCCGGAATGTCACAAAGAAATGTTGGATCGACCCGAGCGC 447
 DB 51 GCTTATTAACCTCTAGGTTCTTCACAGCATGATGATCTGAATCTGGCCAAAGTTC 110

QY 448 GCTAATGTTGATCTTCATCTTACTCAAGAAGAAATGCCGCTGGCTTTGA 503
 DB 111 CATACTGATGACGCTACTGCAATTTAGCACTAATGCAATGCAATGCGCAATTAA 166

RESULT 4
 B2723869 826 bp DNA linear GSS 24-FEB-2003
 LOCUS PUBS097D ZM.0.6.1.0 KB Zea mays genomic clone ZMBMTA097A17,
 DEFINITION genomic survey sequence.
 ACCESSION B2723869
 VERSION B2723869.1 GI:28519534
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 826)
 WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
 Bennetzen,J.
 Maize Genomics Consortium
 Unpublished (2003)
 CONTACT: Cathy Whitelaw
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Classes: sheared ends.
 Location/Qualifiers
 1..826
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZMBMTA097A17"
 /note="Vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
 COT selected genomic DNA library"

ORIGIN

Query Match 5.1%; Score 40.8; DB 28; Length 826;
 Best Local Similarity 59.5%; Pred. No. 0.78;
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 388 GCTTATTAACCCGGAATGTCACAAAGAAATGTTGGATCGACCCGAGCGC 447
 DB 227 GCTTATTAACCTCTAGGTTCTTCACAGCATGATGATCTGAATCTGGCCAAAGTTC 286

QY 448 GCTAATGTTGATCTTCATCTTACTCAAGAAGAAATGCCGCTGGCTTTGA 503
 DB 287 CATACTGATGACGCTACTGCAATTTAGCACTAATGCAATGCGCAATTAA 342

RESULT 5
 CC389867 919 bp DNA linear GSS 19-MAY-2003
 LOCUS PUE1W66TD ZM.0.6.1.0 KB Zea mays genomic clone ZMBMTA262J11,
 DEFINITION genomic survey sequence.
 ACCESSION CC389867

VERSION
 CC389867.1 GI:30869957

KEYWORDS
 GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 919)
 WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
 Bennetzen,J.
 Maize Genomics Consortium
 Unpublished (2003)
 CONTACT: Cathy Whitelaw
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Classes: sheared ends.
 Location/Qualifiers
 1..919
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZMBMTA262J11"
 /note="Vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
 COT selected genomic DNA library"

ORIGIN

Query Match 5.1%; Score 40.8; DB 28; Length 919;
 Best Local Similarity 59.5%; Pred. No. 0.82;
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 388 GCTTATTAACCCGGAATGTCACAAAGAAATGTTGGATCGACCCGAGCGC 447
 DB 506 GCTTATTAACCTCTAGGTTCTTCACAGCATGATGATCTGAATCTGGCCAAAGTTC 565

QY 448 GCTAATGTTGATCTTCATCTTACTCAAGAAGAAATGCCGCTGGCTTTGA 503
 DB 566 CATACTGATGACGCTACTGCAATTTAGCACTAATGCAATGCGCAATTAA 621

RESULT 6
 CF090249 438 bp mRNA linear EST 22-JUL-2003
 LOCUS QHM3P15.Y9.dbl QH M sunflower H. argophyllus Helianthus argophyllus
 DEFINITION cDNA clone QHM3P15, mRNA sequence.
 ACCESSION CF090249
 VERSION CF090249.1 GI:33129316
 KEYWORDS EST.
 SOURCE Helianthus argophyllus
 ORGANISM Helianthus argophyllus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;
 Helianthae; Helianthus.
 1 (bases 1 to 438)
 Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
 Ellison,P., Kolikman,U., Slabough,M.S., Livingston,K., Zhou,Y.,
 Lai,Z., Church,S., Jackson,L. and Bradori,K.
 lettuce and sunflower ESTs from the Compositeae Genome Project
 http://compgenome.ucdavis.edu/
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Armadson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to config QH_CA_Config2435, see http://cspdb.ucdavis.edu/
 for details.
 Plate: QHM12 row: P column: 15.
 Location/Qualifiers

FEATURES

source

1..438
 /organism="Helianthus argophyllus"
 /mol_type="mRNA"
 /db_xref="taxon:73275"
 /clone="QHM3P15"
 /lab_host="E.coli"
 /clone_1lb="QH M sunflower H. argophyllus"
 /note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at http://cspdb.ucdavis.edu/"

ORIGIN

Query Match 5.1%; Score 40.4; DB 14; Length 438;
 Best Local Similarity 47.3%; Pred. No. 0.76;
 Matches 122; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
 QY 451 AATTGTTGATCTTCATCTTACTCAAGAGAAATGCGCTGCTTTGATCATGTC 510
 DB 44 AAATCATATGATACACATCTTAAATCAAGAGATGCGCTGATATGTAACAATC 103
 QY 511 GCGCAATTCAGCATGCGTGGCAAAATTAACGCGCTGCGTTGAATACATCCGTGTA 570
 DB 104 CAACAAGGTAAATCTTCATGCGCAGATAGTTGGGAGGAGGCGCATCTGCTGTAAC 163
 QY 571 CTATTTTCATTAGAAAAGATTTTAGTGGCGGATTTATGATGCCCAATTT 630
 DB 164 ATATTTGAAAAGAAAACAAAGGTTAATGCGCAATTTATGCTAAGAAATATCATC 223
 QY 631 GGCGTCAAGTTAATGCGATGACATTAAGCGCGCTTGATCAATCCGCGTATGTC 690
 DB 224 CTCGAGATATACATTTTGCATAGAGTTCTTGTTGGGTGAACAAAGGGGATTTGTCG 283
 QY 691 AGTACCAATAAATGCGC 708
 DB 284 AGTACCCCTACCGTTGGC 301

RESULT 7

CF087071

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF087071 464 bp mRNA linear EST 22-JUL-2003
 QHM12N04.YG.ab1 QH M sunflower H. argophyllus Helianthus argophyllus
 cDNA clone QHM12N04, mRNA sequence.
 CF087071
 CF087071.1 GI:33126122
 EST.
 Helianthus argophyllus
 Helianthus argophyllus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;
 Helianthaceae; Helianthus.
 1 (bases 1 to 464)
 Kozik, A., Michelmore, R.W., Kapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Davelle, D., Chevalier, P., Ziegler, J.,
 Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositeae Genome Project
 http://compenomics.ucdavis.edu/
 unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Armadson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to config QH_CA_Config2435, see http://cspdb.ucdavis.edu/
 for details.
 Plate: QHM12 row: N column: 04.
 Location/Qualifiers

FEATURES

source

1..464
 /organism="Helianthus argophyllus"
 /mol_type="mRNA"
 /db_xref="taxon:73275"
 /clone="QHM12N04"
 /lab_host="E.coli"
 /clone_1lb="QH M sunflower H. argophyllus"
 /note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at http://cspdb.ucdavis.edu/"

ORIGIN

Query Match 5.1%; Score 40.4; DB 14; Length 464;
 Best Local Similarity 47.3%; Pred. No. 0.78;
 Matches 122; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
 QY 451 AATTGTTGATCTTCATCTTACTCAAGAGAAATGCGCTGCTTTGATCATGTC 510
 DB 53 AAATCATATGATACACACATCTTAAATCAAGAGATGCGCTGATATGTAACAATC 112
 QY 511 GCGCAATTCAGCATGCGTGGCAAAATTAACGCGCTGCGTTGAATACATCCGTGTA 570
 DB 113 CAACAAGGTAAATCTTCATGCGCAGATAGTTGGGAGGAGGCGCATCTGCTGTAAC 172
 QY 571 CTATTTTCATTAGAAAAGATTTTAGTGGCGGATTTATGATGCCCAATTT 630
 DB 173 ATATTTGAAAAGAAAACAAAGGTTAATGCGCAATTTATGCTAAGAAATATCATC 232
 QY 631 GGCGTCAAGTTAATGCGATGACATTAAGCGCGCTTGATCAATCCGCGTATGTC 690
 DB 223 CTCGAGATATACATTTTGCATAGAGTTCTTGTTGGGTGAACAAAGGGGATTTGTCG 292
 QY 691 AGTACCAATAAATGCGC 708
 DB 293 AGTACCCCTACCGTTGGC 310

RESULT 8

AZ396257/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AZ396257 638 bp DNA linear GSS 03-OCT-2000
 M0160C17R Mouse 10kb plasmid UUC1M library Mus musculus genomic
 clone UUC1M0160C17 R, genomic survey sequence.
 AZ396257
 AZ396257.1 GI:10511329
 GSS.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 638)
 Dunn, D., Aoyagi, A., Barber, M., Baecorn, T., Duval, B., Hamill, C.,
 Ismail, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tinsley, A., von
 Niederhausern, A. and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Place: 0168 row C column: 17
 Seq primer: CACACAGGAAACAGCATATACC
 Class: plasmid ends
 High quality sequence stop: 638.
 Location/Qualifiers

FEATURES

source

1. 638
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUC1M0160C17"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_1lb="Mouse 10kb plasmid UUC1M library"
 /note="Vector: PMD42n; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/nares/). The DNA was hydrodynamically sheared by repeated passages through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (gi4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 5.0%; Score 40.2; DB 28; Length 638;
 Best Local Similarity 52.7%; Pred. No. 1;
 Matches 87; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 411 CAACAAATAGAAATTTTGTATGACCCGAGCGGCTAATGGTTGATCTTCAATAC 470
 DB 622 CAGATGAAAGAAATTTTGGCGCTATCCACAACTGCTTACTTAAGAAACAATA 563
 QY 471 TTACTCAAGAGAAATGCGCTGCTTTGATCATGTCCGCAATTCAGCATGCGTG 530
 DB 562 ATTGGTCATTAAGAGAGCAGAGCTTTCTTCTGTGTCGAGAAATTAAGTGCA 503
 QY 531 GCAAAATTAACGCGCTGCGTTGATACATCCCTGTACTATT 575
 DB 502 GAGAAATTAACACCTCATAGAAATCACTCATGTCAGAGATT 458

RESULT 9
 AJ539921 401 bp mRNA linear EST 12-FEB-2003
 LOCUS AJ539921 Hasem3 Helianthus annuus cDNA clone HM007, mRNA

DEFINITION AJ539921 Hasem3 Helianthus annuus cDNA clone HM007, mRNA
 sequence.

ACCESSION AJ539921
 VERSION AJ539921
 KEYWORDS GI:28370224
 EST.
 SOURCE Helianthus annuus (common sunflower)
 ORGANISM Helianthus annuus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;
 Heliantheae; Helianthus.
 1 (bases 1 to 401)
 REFERENCE Tamboirdeguy, C., Laboz, T., Petitprez, M. and Genzbitel, L.
 An expressed-sequence-tag database of the sunflower protoplast
 Unpublished (2003)
 JOURNAL
 COMMENT Contact: Tamboirdeguy C

FEATURES

source

Laboratoire de Biotechnologie et Amélioration des Plantes
 Institut National Polytechnique de Toulouse - Ecole National
 Supérieure Agronomique de Toulouse
 IRR40, Pole de Biotechnologie Vegetale, 18 chemin de Borde Rouge,
 Auzeville, CASTANER TOLOSAN 31326, France.
 Location/Qualifiers
 1. 401
 /organism="Helianthus annuus"
 /mol_type="mRNA"
 /culturivar="Em11"
 /db_xref="taxon:4232"
 /clone="HM007"
 /tissue_type="hypocotyl"
 /cell_type="protoplast"
 /dev_stage="1- to 5-days old protoplast"
 /clone_1lb="Hasem3"

ORIGIN

Query Match 5.0%; Score 39.8; DB 9; Length 401;
 Best Local Similarity 47.4%; Pred. No. 1.1;
 Matches 119; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 458 TTGATCTTCACTTACTTACTCAAGAAATGCGGCTGCTTTGATCATGTCCGCA 517
 DB 33 TAGATCACATTACTTAATACAAAGAGATGCGCTCAGTATGTGAACAAGTCAAC 92
 QY 518 TTGAGATGCGTGGCAAAATTAAGCGGCTGCGGTTGATACATCCGCTACTATT 577
 DB 93 GTAAACTCATGTGCGAGACTTGTGGGAAAGAGAGACTGTGCTTAAATATTTG 152
 QY 578 CATTAGAAAGAGTTTATGTCGCGGATTTATGATGCTTACGCCAATTTGGCGTG 637
 DB 153 GAAAGGAAACACAAAGTTAATGCCCAATTTCTGAGGAACTATCATCATCAGA 212
 QY 638 AAGTTATGCGATGACATTAAGCGGCTGATCAATACCGGGGTGATGACAGCA 697
 DB 213 TTACATTTGCGATGAGATTCTTGTGTGGTGAACAAAGGGGATTGCTAGATCCC 272
 QY 698 ATAAATGGCC 708
 DB 273 CTACTGTGGC 283

RESULT 10
 CF090254 435 bp mRNA linear EST 22-JUL-2003
 LOCUS CF090254 OHM3P22.YG.ab1 OH M sunflower H. argophyllus Helianthus argophyllus
 DEFINITION OHM3P22.YG.ab1 OH M sunflower H. argophyllus Helianthus argophyllus
 cDNA clone OHM3P22, mRNA sequence.
 ACCESSION CF090254
 VERSION CF090254
 KEYWORDS GI:33129321
 EST.
 SOURCE Helianthus argophyllus
 ORGANISM Helianthus argophyllus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;
 Heliantheae; Helianthus.
 1 (bases 1 to 435)
 REFERENCE Kozik, A., Micheltore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Jin, H., van Damme, M., Lavelle, D., Chevallier, P., Ziegler, J.,
 Ellison, P., Kolkmann, D., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Composite Genome Project
 http://comgenomics.ucdavis.edu/
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Micheltore]
 Department of Vegetable Crops, R.W.Micheltore lab
 University of California at Davis (UCD)
 Amundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [micheltore@vegmail.ucdavis.edu]
 belongs to config OH_CA_Config435, see http://cgpdb.ucdavis.edu/

JOURNAL
 COMMENT

for details.

Plate: QHM3 row: P column: 22.

FEATURES

Location/Qualifiers

1..435

/organism="Heliandthus argophyllus"

/mol_type="mRNA"

/db_xref="taxon:73275"

/clone="QHM3P22"

/lab_host="E.coli"

/clone_lib="QH M sunflower H: argophyllus"

/note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at <http://cgpdb.ucdavis.edu/>"

ORIGIN

Query Match

Best Local Similarity 47.4%; Pred. No. 1.2; Length 435;

Matches 119; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 458 TTGATCTTCACTTACTTACTTAAAGAAAGATGCGCTGCTTTGATCATGTGCGGAAA 517
 DB 27 TAGATACACACATCTAATATCAAGAGATGCGCTCATGTATGTGAACAGTCCAAAG 86
 QY 518 TTGAGCATGCGTGGCAAAATTAACGCGCTGCGGTGATACATCCGTGATATTTT 577
 DB 87 GTAAACCTTGTCGCGCAAGTTAGTTGGGAGAGAGCGCATGTGCTGTAAACATATTTG 146
 QY 578 CATTAGAAAAGATTTTATGTCGCGATATTAATGATGCTTACAGCCAAATTTGCGCTGC 637
 DB 147 AAAAGAAAACCAAGAGTTAATGCCCAATTTTCTGAAGAACTATCATCTCTGAGA 206
 QY 638 AAGTTAATGCCAGCATTAAACGCCGCTGTGATCATACCGGGTGATGCTCAGTACCA 697
 DB 207 TATACATTTGCGATAGTTCTTGTGTGGTGAACAAAGGAGATTGTCGTGAGTACC 266
 QY 698 ATAAATGCGC 708
 DB 267 CTACCGTTGCG 277

RESULT 11
 BX500256/c 437 bp mRNA linear EST 04-SHP-2003
 LOCUS DKFZ779J2254_r1_779 (synonym: hnccl) Homo sapiens cDNA clone

DEFINITION DKFZ779J2254_r1_779, mRNA sequence.
 ACCESSION BX500256
 VERSION BX500256.1 GI:32019447
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 437)

Boecher, H., Boecher, M., Mewes, H.W., Weil, B., Amid, C., Oesinger, A., Fodor, G., Han, M. and Wiemann, S.

EST (Boecher, H., Boecher, M., Mewes, H.W., Weil, B., Amid, C., et al.)
 Unpublished (2003)

TITLE JOURNAL
 COMMENT

MPIS
 Ingolstaedter Landstr. 1, D-85764 Neuberg, Germany

This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by GBS (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the
 German Genome Project.

No 5' sequence available.
 This clone (DKFZ779J2254) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Location/Qualifiers

1..437

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZ779J2254"

/issue_type="liver"

/dev_stage="fetal"

/lab_host="DH10B"

/clone_lib="779 (synonym: hnccl)"
 /note="Vector: pSPORT1_Sfi, Site_1: SfiIA, Site_2: SfiIB"

ORIGIN

Query Match

Best Local Similarity 53.2%; Pred. No. 1.3; Length 437;

Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 237 TGGAGAACCTGTGATGAGATTTTGTATGACGAGACGATTTTGTGCGGAAAAGT 296
 DB 275 TCGAGACCTTTCCATTCATTTTAAAGCTGTGAGATTTTTCATTAAGAAAGT 216
 QY 297 CCATCTTATCCATTTTTCAGCCATCCGCGTTGATGAGCAACCCAAAGCGATCC 356
 DB 215 CTGCTTTCAAGTGAACATCATGCTCCGCTTTGTGTGCTTAAATGATTTGT 156
 QY 357 GAATGTTGAGATGTCAGTATTTCCATTTACGCTTTAT 394
 DB 155 GACTTTTTCAGAGACCTTCATTTCTGTACATTTGAT 118

RESULT 12
 CF091479 447 bp mRNA linear EST 22-JUL-2003
 LOCUS OHM8P22.Y9.ab1.QH M sunflower H: argophyllus Heliandthus argophyllus

DEFINITION OHM8P22.Y9.ab1.QH M sunflower H: argophyllus Heliandthus argophyllus
 cDNA clone QHM8P22, mRNA sequence.

ACCESSION CF091479
 VERSION CF091479.1 GI:33130546
 KEYWORDS EST.

SOURCE Heliandthus argophyllus
 ORGANISM Heliandthus argophyllus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asterales; Asteraceae; Asteraceae; Heliandthus.

1 (bases 1 to 447)
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,

Lin, H., Van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,

Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,

Lai, Z., Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositeae Genome Project

<http://compenomics.ucdavis.edu/>
 Unpublished (2002)

Contact: Alexander Kozik (R.W. Michelmore)
 Department of Vegetable Crops, R.W. Michelmore Lab

University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmorevegmail.ucdavis.edu]
 belongs to contig QH_CA_Contig2435_2, see <http://cgpdb.ucdavis.edu/>

for details.
 Plate: QHM8 row: F column: 22.

FEATURES
 source Location/Qualifiers

1..447
 /organism="Heliandthus argophyllus"
 /mol_type="mRNA"
 /db_xref="taxon:73275"
 /clone="QHM8P22"
 /lab_host="E.coli"
 /clone_lib="QH M sunflower H: argophyllus"
 /note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling root and leaf) of RNA from a single genotype. cDNAs were pooled and

ORIGIN

directionally cloned into a custom medium-copy vector.
Details of library construction can be obtained at
<http://cspdb.ucdavis.edu/>

Query Match 4.9%; Score 39.4; DB 14; Length 447;
Best Local Similarity 47.4%; Pred. No. 1.6;
Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 460 GATCTTCATCTTCTACTCAAGAGAAATGCGCTGCTTTTGATCATGTCCGCAAT 519
DB 21 GATACACACATCTTAATACAAAGAGATGGCTTCAGATGTGAACAGATCAAGGT 80
QY 520 CAGCATGCGTGGCAAAATTAACGCGTGGTGAATACATCCGTGCTACTATTTC 579
DB 81 AAACTTCGTGCGCAAGCTAGTCCGGAGAGAGATTTGCTGTAAACATATTGA 140
QY 580 TTAGAAAAAGATTTTAAAGTGGCGGATATTATGATGCTTACGCCAAATTTGGCTGAA 639
DB 141 AAGGAAAACGCGAAGGTTAATGCGCAAACTATCGCAAGAACACGATCCCGAGATA 200
QY 640 GTTAATCGATGACCATTAAGCGCGCTGATCAATACCGGGGTGATGCTGATACCAAT 699
DB 201 TACATTTGCGATGAGTCTTGTGTGGTGAACAAAGAGGAGATTGTGTGATGATCCCT 260
QY 700 AAAATGACC 708
DB 261 ACCGTTGGC 269

RESULT 13

CF090121 454 bp mRNA linear EST 22-JUL-2003
LOCUS QHMJ13.YG.ab1 OH M sunflower H. argophyllus Helianthus argophyllus
DEFINITION cDNA clone QHMJ13, mRNA sequence.
ACCESSION CF090121
VERSION CF090121.1 GI:33129188

SOURCE

EST:
Helianthus argophyllus
Helianthus argophyllus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Asteroideae;
Heliantheae; Helianthus.

REFERENCE

1 (bases 1 to 454)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositeae Genome Project
<http://compgenomics.ucdavis.edu/>
Unpublished (2002)

JOURNAL

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Aamundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OH_CA_Config2435_2, see <http://cspdb.ucdavis.edu/>
for details.
Plate: QHM3 row: J column: 13.

FEATURES

source

1..454
Location/Qualifiers
/organism="Helianthus argophyllus"
/mol_type="mRNA"
/db_xref="taxon:73275"
/clone="QHMJ13"
/lab_host="E.coli"
/clone_lib="OH M sunflower H. argophyllus"
/note="Vector: pBRCDNA5f1AB. The library was constructed
from three different sources (seedling, root and leaf) of
RNA from a single genotype. cDNAs were pooled and

ORIGIN

directionally cloned into a custom medium-copy vector.
Details of library construction can be obtained at
<http://cspdb.ucdavis.edu/>

Query Match 4.9%; Score 39.4; DB 14; Length 454;
Best Local Similarity 47.4%; Pred. No. 1.6;
Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 460 GATCTTCATCTTCTACTCAAGAGAAATGCGCTGCTTTTGATCATGTCCGCAAT 519
DB 57 GATACACACATCTTAATACAAAGAGATGGCTTCAGATGTGAACAGATCAAGGT 116
QY 520 CAGCATGCGTGGCAAAATTAACGCGTGGTGAATACATCCGTGCTACTATTTC 579
DB 117 AAACTTCGTGCGCAAGCTAGTCCGGAGAGAGATTTGCTGTAAACATATTGA 176
QY 580 TTAGAAAAAGATTTTAAAGTGGCGGATATTATGATGCTTACGCCAAATTTGGCTGAA 639
DB 177 AAGGAAAACGCGAAGGTTAATGCGCAAACTATCGCAAGAACACGATCCCGAGATA 236
QY 640 GTTAATCGATGACCATTAAGCGCGCTGATCAATACCGGGGTGATGCTGATACCAAT 699
DB 237 TACATTTGCGATGAGTCTTGTGTGGTGAACAAAGAGGAGATTGTGTGATGATCCCT 296
QY 700 AAAATGACC 708
DB 297 ACCGTTGAC 305

RESULT 14

CF096847 401 bp mRNA linear EST 22-JUL-2003
LOCUS QHN2F11.YG.ab1 OH N sunflower H. argophyllus (drought stress)
DEFINITION Helianthus argophyllus cDNA clone QHN2F11, mRNA sequence.
ACCESSION CF096847
VERSION CF096847.1 GI:33135914

SOURCE

EST:
Helianthus argophyllus
Helianthus argophyllus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Asteroideae;
Heliantheae; Helianthus.

REFERENCE

1 (bases 1 to 401)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositeae Genome Project
<http://compgenomics.ucdavis.edu/>
Unpublished (2002)

JOURNAL

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Aamundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OH_CA_Config2435_2, see <http://cspdb.ucdavis.edu/>
for details.
Plate: QHN2 row: F column: 11.

FEATURES

source

1..401
Location/Qualifiers
/organism="Helianthus argophyllus"
/mol_type="mRNA"
/db_xref="taxon:73275"
/clone="QHN2F11"
/lab_host="E.coli"
/clone_lib="OH N sunflower H. argophyllus (drought stress)"
/note="Vector: pGEM-T. The library was constructed from
three different sources (seedling, root and leaf) of RNA
from a single genotype. cDNAs were pooled and cloned into

ORIGIN

a high-copy vector pGBM-T. Details of library construction can be obtained at <http://cspdb.ucdavis.edu/>

Query Match 4.9%; Score 39; DB 14; Length 401;
Best Local Similarity 46.5%; Pred. No. 1.9;
Matches 126; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 455 GGTGATCTTCACTTACTTACTCAAGAAGAAATGCCCGCTTGTGATCATGTGCGC 514
DB 5 GGATGACACACACATCTTAATCAAGAGATGCTCATGTATGAACAAGTCCAC 64
QY 515 AATTGAGATGCGTGGCAAAATTAACGCGCTGATGATCAATCCGCTGACTAT 574
DB 65 AAGTAAACTCTGCGCCAGAGCTGTCGGGAAAGAGAGATGTGCTGTAACATTA 124
QY 575 TTTCATTAGAAAAGAGTTTAAAGTGGCGGATTTATGATGCTACGCCAATTGCGC 634
DB 125 TTGAAGAAGAAACGGAGGTTAATGCCCAACTATGCGAGAGAAACGACATCCCG 184
QY 635 TCGAAGTTAATCGCATGACATTAAACGCCGCTGATCAATACCGGGGATTCGTA 694
DB 185 AGATATACATTTGCGATAGATTTCTGTGTGGTGAACAGAGGGGATGTGTA 244
QY 695 CCAATTAATGCGCCGATCTTGTAAAGCAA 725
DB 245 CCCCTACGCTTGCTAATAAAAAAAAAACA 275

RESULT 15

AJ540178 451 bp mRNA linear EST 12-FEB-2003
AJ540178 Hasems3 Helianthus annuus cDNA clone HM0013, mRNA
LOCUS
DEFINITION
sequence.

ACCESSION AJ540178
VERSION AJ540178.1 GI:28370481
KEYWORDS EST
SOURCE Helianthus annuus (common sunflower)
ORGANISM Helianthus annuus

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Tamboirindeguy, C., Liboz, T., Petitprez, M. and Gentzbitel, L.
JOURNAL An expressed-sequence-tag database of the sunflower protoplast
COMMENT Unpublished (2003)
Contact: Tamboirindeguy C
Laboratoire de Biotechnologie et Amélioration des Plantes
Institut National Polytechnique de Toulouse - Ecole National
Supérieure Agronomique de Toulouse
IFR40, Pole de Biotechnologie Vegetale, 18 chemin de Borde Rouge,
Auzeville, CASTANET TOLOSAN 31326, France.

FEATURES

source
1..451
Location/Qualifiers
/organism="Helianthus annuus"
/mol_type="mRNA"
/cultivar="Emil"
/db_xref="taxon:4232"
/clone="HM0013"
/issue_type="hyprocotyl"
/cell_type="protoplast"
/dev_stage="1- to 5-days old protoplast"
/clone_lib="Hasems3"

ORIGIN

Query Match 4.9%; Score 39; DB 9; Length 451;
Best Local Similarity 47.7%; Pred. No. 2.1;
Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 458 TTGATCTTCACTTACTTACTCAAGAAGAAATGCCGCTTGTGATCATGTGCGCAA 517
DB 32 TAGATACACACATCTAATCAACAAGAGGTGCCCCAGTATGTGAACAAGTCAACAAG 91

QY 518 TTCAGCATGCGTGGCAAAATTAACGCGCTGCGGTTGATATACATCCGTAATTTT 577
DB 92 GTAAACTTGTGCGCAGAGCTAGTGGGGAAGAGATTGTGTGTAACATATATG 151
QY 578 CATTAGAAAAGAGTTTAAAGTGGCGGATTTATGATGCTTACGCCAATTGCGCTG 637
DB 152 AAAAGAAAACGGAGGTTAATGCCCAACTATGCGAAGAGAACACGATCCCGAGA 211
QY 638 AAGTTAATCGCATGACCATTAACGCCGCTTGTATCATACCGGGGTGATGTCATAC 696
DB 212 TATACATTTGCGATAGATTTCTGTGTGGGTGAACAAAGGGGATTGTGTGATAC 270

Search completed: April 30, 2004, 18:35:03
Job time : 2273.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:32 ; Search time 93.7866 Seconds

(without alignments) 4721.887 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798
Sequence: 1 acgttcgcacaaagctcta.....ttcaactgtttacagagt 798

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| 1 | 35.6 | 4.5 | 552 | 4 | US-09-107-532A-286 |
| 2 | 35.2 | 4.4 | 4392 | 1 | US-08-026-138E-5 |
| 3 | 34 | 4.3 | 1326 | 4 | US-09-489-039A-2920 |
| 4 | 33.6 | 4.2 | 2628 | 4 | US-09-294-531B-5 |
| 5 | 33.4 | 4.2 | 474 | 4 | US-09-621-976-18033 |
| 6 | 33.4 | 4.1 | 1830121 | 4 | US-09-557-884-1 |
| 7 | 33 | 4.1 | 1830121 | 4 | US-09-643-990A-1 |
| 8 | 32.6 | 4.1 | 552 | 4 | US-09-489-039A-10 |
| 9 | 32.6 | 4.1 | 825 | 4 | US-09-134-001C-2442 |
| 10 | 32.4 | 4.1 | 696 | 3 | US-08-998-41E-1134 |
| 11 | 32.4 | 4.1 | 1620 | 3 | US-08-925-230-2 |
| 12 | 32.4 | 4.1 | 1620 | 4 | US-09-712-372-2 |
| 13 | 32.4 | 4.1 | 11466 | 4 | US-08-956-171E-444 |
| 14 | 32.4 | 4.0 | 846 | 4 | US-09-328-352-4044 |
| 15 | 32 | 4.0 | 7493 | 1 | US-08-212-133A-7 |
| 16 | 32 | 4.0 | 7493 | 1 | US-08-474-503-5 |
| 17 | 32 | 4.0 | 7493 | 2 | US-08-670-707A-5 |
| 18 | 32 | 4.0 | 7493 | 3 | US-09-037-601-5 |
| 19 | 32 | 4.0 | 7493 | 3 | US-09-315-179-5 |
| 20 | 32 | 4.0 | 7493 | 5 | PCT-US94-13200-5 |
| 21 | 31.6 | 4.0 | 378 | 3 | US-09-328-11-191 |
| 22 | 31.6 | 4.0 | 1446 | 2 | US-08-933-750C-91 |
| 23 | 31.6 | 4.0 | 1724 | 3 | US-09-234-613-91 |
| 24 | 31.6 | 4.0 | 1724 | 3 | US-09-620-312D-650 |
| 25 | 31.4 | 3.9 | 1321 | 4 | US-09-620-312D-790 |
| 26 | 31.4 | 3.9 | 1462 | 4 | US-09-620-312D-788 |
| 27 | 31.4 | 3.9 | 1519 | 4 | US-09-620-312D-789 |

| | | | | | | |
|----|------|-----|--------|---|---------------------|--------------------|
| 28 | 31.4 | 3.9 | 2758 | 4 | US-09-221-017B-441 | Sequence 441, App |
| 29 | 31.2 | 3.9 | 2290 | 4 | US-09-620-312D-891 | Sequence 891, App |
| 30 | 31 | 3.9 | 477 | 4 | US-09-489-039A-801 | Sequence 801, App |
| 31 | 31 | 3.9 | 1089 | 4 | US-09-134-000C-901 | Sequence 901, App |
| 32 | 30.8 | 3.9 | 648 | 4 | US-09-543-681A-3313 | Sequence 3313, App |
| 33 | 30.8 | 3.9 | 1152 | 4 | US-09-540-236-1145 | Sequence 1145, App |
| 34 | 30.8 | 3.9 | 119211 | 4 | US-09-596-002-40 | Sequence 40, App |
| 35 | 30.6 | 3.8 | 500 | 3 | US-09-141-000-2 | Sequence 2, App |
| 36 | 30.6 | 3.8 | 678 | 3 | US-09-232-479-19 | Sequence 19, App |
| 37 | 30.6 | 3.8 | 678 | 4 | US-09-784-990-19 | Sequence 19, App |
| 38 | 30.6 | 3.8 | 754 | 4 | US-09-221-017B-941 | Sequence 941, App |
| 39 | 30.6 | 3.8 | 2829 | 3 | US-08-851-843A-53 | Sequence 53, App |
| 40 | 30.6 | 3.8 | 2829 | 3 | US-08-974-549A-220 | Sequence 220, App |
| 41 | 30.6 | 3.8 | 2829 | 3 | US-08-854-050-53 | Sequence 53, App |
| 42 | 30.6 | 3.8 | 2829 | 4 | US-09-430-123-53 | Sequence 53, App |
| 43 | 30.6 | 3.8 | 2829 | 4 | US-09-402-181B-220 | Sequence 220, App |
| 44 | 30.6 | 3.8 | 2829 | 4 | US-09-721-456-220 | Sequence 220, App |
| 45 | 30.6 | 3.8 | 3300 | 4 | US-09-268-347-31 | Sequence 31, App |

ALIGNMENTS

RESULT 1
US-09-107-532A-286
Sequence 286, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSER: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Denek

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 286:

SEQUENCE CHARACTERISTICS:

LENGTH: 552 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

MOLECULE TYPE: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc.feature

LOCATION: (B) LOCATION 1...552


```

1 SOFTWARE: FastSeq for Windows Version 3.0
2 SEQ ID NO 5
3 LENGTH: 2628
4
5 TYPE: DNA
6 ORGANISM: Artificial Sequence
7 FEATURE:
8 OTHER INFORMATION: Degenerate nucleotide sequence encoding the
9 OTHER INFORMATION: polypeptide of SEQ ID NO:2
10 NAME/KEY: variation
11 LOCATION: (1)...(2628)
12 OTHER INFORMATION: Each N is independently any one of A, T, G or C
13
14 US-09-294,531B-5

```

| | | | | |
|-----------------------|------------------|-----------------|-----------|--------------|
| Query Match | 4.2%; | Score 33.6; | DB 4; | Length 2628; |
| Best Local Similarity | 32.6%; | Pred. No. 0.62; | | |
| Matches 45; | Conservative 27; | Mismatches 66; | Indels 0; | Gaps 0; |

Dy 628 TTGGCGTCGAAGTTATCGATGACATTAAACGCCGCTTATCATACCGGGGTATC 687
Db 894 YTTNGCGNCNACRTGCATNACRTTPRANCKKCANCCYTCYNACDATINGTGRTPTC 835

Qy 688 GTCA GTACCAATAAATGCGCCATCTTGTAAAGGCAAGGACCAACCAAGCACCGTT 747

Dib 834 NCCSMWCCDATNARRTANCXGGRITNSWCCADATNARRTCNSWRANGRTCTCTCTC 775

| | | | |
|----|-----|--------------------|-----|
| QY | 748 | TATCGTCTTGCCAGTCAT | 765 |
| | | : : | : |
| Db | 774 | NCCYTTYTCNACNGCYTT | 757 |

RESULT 5
US-09-621-976-18033
: Sequence 18033. Application US/09621976

```

; Patent NO. 6639063
;
; GENERAL INFORMATION:
;
; APPLICANT: Dumas Milne Edwards, J.B.
;
; APPLICANT: Robert, S.

```

APPLICANT: Giordano, U.I.
TITLE OF INVENTION: ESTs and Encoded Human Proteins
FILE REFERENCE: GENSET:054PR2
CURRENT APPLICATION NUMBER: US/09/621 976

```

; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent .pm
; SEQ ID NO 18033

```

```

; LENGTH: 373
;
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

US-09-621-976-18033

| | | | | |
|-----------------------|-------------------|-----------------|-----------|-------------|
| Query Match | 4.2%; | Score 33.4; | DB 4; | Length 474; |
| Best Local Similarity | 12.5%; | Pred. NO. 0.25; | | |
| Matches 34; | Conservative 118; | Mismatches 119; | Indels 0; | Gaps 0; |

Dy 528 GTGCGAAAAATTACGGCTGCGGTGAATACACATCCGTGGTACTAATTTCA TTAGAAA 5877

Db 102 KTSCMAGRWKGYYSRMTYYCYICACYIMMKRTVCSGCCMMTYTGCGGSMMTTTTMMRR 161

Oy 588 AGAGTTTAAAGTGCGGATATTATGATGCTACGCCAANAATTGGCGTCGAAGTAATCG 647
:: :: : :: : :: : :: : :: : :: : :: : :: :
Db 162 RKSSYKRWTKGKKKKTTNNNAAMCYTTWASYNMNNRRRAAAAKTYTYCNMSKTMCMAC 221

Dy 648 CATGACCATTTAAACGCCCGCTGATCATATACGGGGGATCGTCAGTACCATTAAATGGC 707

Dd 222 CCMCCMRARSCWRSCKRSTYNNCYYYMMYKSGRMVMMWRGGNMKRMTMYKKKSM 281

Dy 708 CGCATCTTTGTAAGCCAAAGAGCCCAACCGCACGGTATTCGTCTTGCCAGTCAATGA 767

Dd 282 WKSGCMKKAAWMAARKTTYTWAMYTTYTYKMCCTYMKRTTYCMMMYSRMRGSMWTAR 341

```
Oy      768 AGTCACCTATTTCAAACTGTTTACGAGGT   798  
         :|::: |: |:  
Db     342 GANNVCYWMYYMAARKKKKMMWAAAAARGCN   372
```

RESULT 6
US-09-557-884-1
; Sequence 1, Application US/09557884

```

; GENERAL INFORMATION:
;
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
;

```

Thereof, and Uses There
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences Inc

STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA

```

;
;      ZIP: 20850
;
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: 3 1/2 inch diskette
;      COMPUTER: Dell Pentium

```

```

;
; OPERATING SYSTEM: MS DOS V6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: IIS/09/557

```

FILING DATE: 25-APR-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/476,102

FILED DATE: JUN-5-1987
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Mark
REGISTRATION NUMBER: 4

REFERENCE DOCUMENT NUMBER: E11001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid

```

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

| | | | | | | | |
|-----------------------|--------|--------------|-----|------------|-----|--------|----------|
| Query Match | 4.1%; | Score | 33; | DB | 4; | Length | 1830121; |
| Best Local Similarity | 65.8%; | Pred. NO. | 52; | | | | |
| Matches | 48; | Conservative | 0; | Mismatches | 25; | Indels | 0; |
| | | | | | | Gaps | 0; |

| Qy | 586 | AAAGAGTTTAAAGTGGCGGAATTAATTTGATGCTACGGCCAAATTTGCGGTGGAAGTTAAT | 645 |
|----|---------|---|---------|
| Db | 1807122 | AACGATTTTGTGTCGCGTCACTTCAGGCCGTCGCCAATTTCCTTGATTTAAC | 1807181 |

```

Qy      646 CGCATGACCATT 658
          | | | | |
Db      1807182 AGTAATGCCATCA 1807194

```

RESULT 7
US-09-643-990A-1
Sequence 1: Application US/09643990A

Patent NO. 6548489
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams

Owen, Willie
Hamilton O. Smith

J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Query Match 4.1%; Score 33; DB 4; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 52;
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
CY 586 AAGAGTTTATGAGCGGATATTGATGCTAGCCAAATTTGGCGTGAAGTTAT 645
DB 1807122 AACGATTTTGTGCGCGTCAATCTTCAAGCCGTCACACTTTGCTTCGATTTAAC 1807181
CY 646 CGCATGACCATTA 658
DB 1807182 AGTAATCCATCA 1807194
RESULT 8
US-09-489-039A-10/c
Sequence 10, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 10
LENGTH: 552
TYPE: DNA
ORGANISM: Klebsiella pneumoniae

US-09-489-039A-10
Query Match 4.1%; Score 32.6; DB 4; Length 552;
Best Local Similarity 57.3%; Pred. No. 0.52;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
CY 262 GATGACCGGAGACGACGATTTTGTGGCAAAAGTCATCTTACTTATTTTACAC 321
DB 175 GTTTCAGCCCAACACGATGTGTAGTAAAGCCAGACAGATGCTTACGACTTGAACGAC 116
CY 322 GATCCGCTGTTGATGGCAACCCCAACCGGATCCGATGTT 364
DB 115 GATGACCAATTCTACTTACCGCAACGCGCTCGGATATT 73
RESULT 9
US-09-134-001C-2442/c
Sequence 2442, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2442
LENGTH: 825
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2442
Query Match 4.1%; Score 32.6; DB 4; Length 825;
Best Local Similarity 57.3%; Pred. No. 0.67;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
CY 65 TGTATTATGAGAGCCATATGATTTGCTACTATTCATATGCTGCGCTTGGC 124
DB 588 TATTAATTCGAAATTAATGTTGCTGCTTCAATTTCTACTAAGTA 529
CY 125 CGAATCTATCCAGTTTACTGCTGAACGCTCTATCCAA 167
DB 528 TCGAATGTTCCAGCTAATACGCTGTGATGATATCAACAGA 486
RESULT 10
US-08-998-416-1134/c
Sequence 1134, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1134:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1690UP
US-08-998-416-1134

Query Match 4.1%; Score 32.4; DB 3; Length 696;
Best Local Similarity 47.5%; Pred. No. 0.71;
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 171 GCGAGATTGTGTAATGGCGATTGCTGGCGGATAGTATGACGAAGATATACCGC 230
DB 452 GCGAATATGCTGCTGGCGAAGTACGACAGTATTTGCACTAGAGGTGCTGGGATCA 393
QY 231 TCAATGTGAGAACTGTGATGAGATTGATGACGAGACGATATTGTCGCA 290
DB 392 ACTTAGGGGAGACGACTTCGAGAGGATAGAGCGCAAGCGGGGCTAGCGCGAG 333
QY 291 AAAAGCCCTACTTATTCCTAATTTATATGCGATCCGCTGTTGATGCAACCCCAAG 350
DB 332 CAAGGGCGAAACGATCAAGATATAGTGTGCGAGCGGAGCTAGAGTACGACCA 273
QY 351 CGATCCGATGTTGAGTGTGTC 372
DB 272 CGTGAAGATGAGGCGCGTC 251

RESULT 11
US-08-925-230-2
Sequence 2, Application US/08925230
Patent No. 6147194
GENERAL INFORMATION:
APPLICANT: Collart, Frank
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO IMPDH
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/925,230
FILING DATE: September 8, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 5,665,583
FILING DATE: 12-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:274
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1620 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-925-230-2

Query Match 4.1%; Score 32.4; DB 3; Length 1620;
Best Local Similarity 58.2%; Pred. No. 1.2;
Matches 57; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 598 GTGGCGATATTATTGATGCTACGCCAATTGGCGTCCAGATTATCGCATGACCAT 657
DB 404 GTGAGGATGTTTGTGAGCAAGCAAGCGCATGCTCTGTGTATCCCATCAGCAT 463
QY 658 AAACGCCCTTGATCAATACCGGGTGATGTCAGTAC 695
DB 464 ACAAGCCGATGGGAGTGCATGCTGCGGATCATTTTC 501

RESULT 12
US-08-712-372-2
Sequence 2, Application US/09712372
Patent No. 6479628
GENERAL INFORMATION:
APPLICANT: Collart, Frank
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO IMPDH
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/712,372
FILING DATE: 13-NO. 6479628-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/925,230
FILING DATE: September 8, 1997
APPLICATION NUMBER: US 5,665,583
FILING DATE: 12-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:274
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 474-7577
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 1620 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-712-372-2

Query Match 4.1%; Score 32.4; DB 4; Length 1620;
Best Local Similarity 58.2%; Pred. No. 1.2;
Matches 57; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 598 GTGGCGATATTTATGATGCTTACGCGCAATTTGGCGTGAAGTTATGCGATGACCAT 657
DB 404 GTAGGAGATCTTTTGAAGCAAGCCAGGATGCTTGTGATATCCCATCAAGAT 463
QY 658 AAAGCGCGCTTGATCAATACCGGGGTGATGTCATGAC 695
DB 464 ACAGCGCGATGGAGTGCAGTGTGGCGCATCATTTTC 501

RESULT 13

US-08-956-171E-444
Sequence 444, Application US/08956171E
Patent No. 6593114

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 444:

SEQUENCE CHARACTERISTICS:

LENGTH: 1146 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 444:

US-08-956-171E-444

Query Match 4.1%; Score 32.4; DB 4; Length 1146;

Best Local Similarity 60.0%; Pred. No. 4;

Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATGTTTGTCACAAAGTCTTATTACATTGACCATCATGCACTTACTTAATAAGC 60
DB 11258 ATGTTGACTATGACCTTAATTTCTTGTGACCTATATCTTACTTAACAAATATGCTTA 11317
QY 61 CCGTTGTTATTAGGAGGACCATTAAGATT 90
DB 11318 TCGTTGATTAGTAAGTATCAATTTTGGTT 11347

RESULT 14

US-09-328-352-4044/C

Sequence 4044, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 4044

LENGTH: 846

TYPE: DNA

ORGANISM: Acinetobacter baumannii

US-09-328-352-4044

Query Match 4.0%; Score 32; DB 4; Length 846;
Best Local Similarity 47.9%; Pred. No. 1.1;
Matches 92; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 367 AGTGTGATATTTTCCATTAAGCTTTTAAACCCGTGAATGTCAACAAATAGAGAT 426
DB 601 AGCTTAATTTTATACCTGACGACGATCATCTTTCATGAGTGAAGAAATAGCTGAC 542
QY 427 TTGTGATCGAACCCGCGCGCTAATGTTGATTTTCACTTTACTCAAGAGAA 486
DB 541 CTGAACCTGAACCCGCGCTTACGCGATCAATTTGAGTTGAGTTTACCCAAATTGTTA 482
QY 487 ATGCGCGTGGCTTTGATCAATGCGCAATTCAGATCGTGGCAAAATTAACGGCT 546
DB 481 AACAGTTTCTGCTTAATATCTGACTGTTAATTCAGCTCCGAGGACCAACGACATCAG 422
QY 547 GCGGTTGAATAC 558
DB 421 TCGCGTAATATAC 410

RESULT 15

US-08-212-133A-7/C

Sequence 7, Application US/08212133A

Patent No. 5663060

GENERAL INFORMATION:

APPLICANT: Lollar, John S.

TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kiplattick & Cody

STREET: 100 Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: US

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/212,133A

FILING DATE: March 11, 1994

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="5'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product="Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschler, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 563060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

Query Match          4.0%; Score 32; DB 1; Length 7493;
Best Local Similarity 62.5%; Pred No. 4.2;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY      35 TCATCATGCACTTACTTAATAATAGCCCGTGTATTATAGGAGACCATATGATGTCA 94
DB      895 TCTTTCAGAGCTTGCACAACTATAGATGACTTTCACACAGGGAACCTTATCATCTTCC 836
QY      95 CTATGATATGATTTGTCTG 114
DB      835 TTCCTCATTTGGCTTGTCTG 816
```

Search completed: April 30, 2004, 18:43:08
Job time : 108.787 secs

Db 121 GGCCGGAATCTATCCAGGTTTACTGTGTAACGCTCTAATCCAAATCGCGCATTTGT 180
QY 181 GGTAAATGGGCAATTCCTGCGCGGAGTATGATGACGAAGATATGACCGCTCATGTGGA 240
Db 181 GGTAAATGGGCAATTCCTGCGCGGAGTATGATGACGAAGATATGACCGCTCATGTGGA 240
QY 241 GAACCTGTGATGAGGATTTTATGATGACGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300
Db 241 GAACCTGTGATGAGGATTTTATGATGACGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300
QY 301 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360
Db 301 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360
QY 361 GGTGAGTGTCAATTTTCCCATTAAGCTTTATTTAAACCGGTGGAATGTCAAATA 420
Db 361 GGTGAGTGTCAATTTTCCCATTAAGCTTTATTTAAACCGGTGGAATGTCAAATA 420
QY 421 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGTTGATCTTCACTTACTCAA 480
Db 421 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGTTGATCTTCACTTACTCAA 480
QY 481 GAAGAAATCCCGTGGCTTTTATATCATGTGCGCAAAATTCAGCATGCGTGGCAAAATA 540
Db 481 GAAGAAATCCCGTGGCTTTTATATCATGTGCGCAAAATTCAGCATGCGTGGCAAAATA 540
QY 541 CGGCGTGGGTTGATATACATCCGCTGTGATCTTCACTTCAATTAAGAAAGATTTTATG 600
Db 541 CGGCGTGGGTTGATATACATCCGCTGTGATCTTCACTTCAATTAAGAAAGATTTTATG 600
QY 601 GCGGATATTATGATGCTTACGCAAAATTTGCGGCAATTTGATGATGATGATGATGAT 660
Db 601 GCGGATATTATGATGCTTACGCAAAATTTGCGGCAATTTGATGATGATGATGATGAT 660
QY 661 CGCGCTTATATATATCCGCGGTGATGCTGATACCAATTAATGCGGCAATTTGATGAT 720
Db 661 CGCGCTTATATATATCCGCGGTGATGCTGATACCAATTAATGCGGCAATTTGATGAT 720
QY 721 GCGAAGAGAGCAACCAACGCTTATGCTTTCGATGATGATGATGATGATGATGATGAT 780
Db 721 GCGAAGAGAGCAACCAACGCTTATGCTTTCGATGATGATGATGATGATGATGATGAT 780
QY 781 CAAACCTGTTTACGAGGT 798
Db 781 CAAACCTGTTTACGAGGT 798

RESULT 2
US-10-603-260-1
Sequence 1, Application US/10603260
Publication No. US20040009570A1
GENERAL INFORMATION:
APPLICANT: CJ Corporation
TITLE OF INVENTION: An alkaline lipase from *Vibrio metschnikovii* RH530 and a
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/603,260
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Koparentin 1.71
SEQ ID NO 1
LENGTH: 2578
TYPE: DNA
ORGANISM: *Vibrio metschnikovii* RH530
US-10-603-260-1

Query Match 100.0%; Score 798; DB 16; Length 2578;
Best Local Similarity 100.0%; Pred. No. 2,9e-239;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTGTCAAAAGCTTATTTATCACTTGGACCATATATGACCTTACCTTAAATAAGC 60
Db 703 ATGTTTGTCAAAAGCTTATTTATCACTTGGACCATATATGACCTTACCTTAAATAAGC 762

QY 61 CCGTGTGTTTATAGGAGCCATATATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 763 CCGTGTGTTTATAGGAGCCATATATGATGATGATGATGATGATGATGATGATGATGAT 822
QY 121 GCGCGGAATCTATCCAGGTTTATGATGAAACGCTTATATCAATGAGGCAATGT 180
Db 823 GCGCGGAATCTATCCAGGTTTATGATGAAACGCTTATATCAATGAGGCAATGT 882
QY 181 GGTAAATGGGCAATTCCTGCGCGGAGTATGATGACGAAGATATGACCGCTCATGTGGA 240
Db 883 GGTAAATGGGCAATTCCTGCGCGGAGTATGATGACGAAGATATGACCGCTCATGTGGA 240
QY 241 GAACCTGTGATGAGGATTTTATGATGACGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300
Db 943 GAACCTGTGATGAGGATTTTATGATGACGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300
QY 301 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360
Db 1003 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360
QY 361 GGTGAGTGTCAATTTTCCCATTAAGCTTTATTTAAACCGGTGGAATGTCAAATA 420
Db 1063 GGTGAGTGTCAATTTTCCCATTAAGCTTTATTTAAACCGGTGGAATGTCAAATA 420
QY 421 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGTTGATCTTCACTTACTCAA 480
Db 1123 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGTTGATCTTCACTTACTCAA 480
QY 481 GAAGAAATCCCGTGGCTTTTATATCATGTGCGCAAAATTCAGCATGCGTGGCAAAATA 540
Db 1183 GAAGAAATCCCGTGGCTTTTATATCATGTGCGCAAAATTCAGCATGCGTGGCAAAATA 540
QY 541 CGGCGTGGGTTGATATACATCCGCTGTGATCTTCACTTCAATTAAGAAAGATTTTATG 600
Db 1243 CGGCGTGGGTTGATATACATCCGCTGTGATCTTCACTTCAATTAAGAAAGATTTTATG 600
QY 601 GCGGATATTATGATGCTTACGCAAAATTTGCGGCAATTTGATGATGATGATGATGATGAT 660
Db 1303 GCGGATATTATGATGCTTACGCAAAATTTGCGGCAATTTGATGATGATGATGATGATGAT 660
QY 661 CGCGCTTATATATATCCGCGGTGATGCTGATACCAATTAATGCGGCAATTTGATGAT 720
Db 1363 CGCGCTTATATATATCCGCGGTGATGCTGATACCAATTAATGCGGCAATTTGATGAT 720
QY 721 GCGAAGAGAGCAACCAACGCTTATGCTTTCGATGATGATGATGATGATGATGATGATGAT 780
Db 1423 GCGAAGAGAGCAACCAACGCTTATGCTTTCGATGATGATGATGATGATGATGATGATGAT 780
QY 781 CAAACCTGTTTACGAGGT 798
Db 1483 CAAACCTGTTTACGAGGT 1500

RESULT 3
US-10-282-122A-41398
Sequence 41398, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zysek, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EUTRA.034A

;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P330R1C24
;; CURRENT APPLICATION NUMBER: US/10/142,426
;; CURRENT FILING DATE: 2002-05-09
;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 426
;; LENGTH: 747
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-142-426-426

Query Match 4.4%; Score 35; DB 13; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5; Mismatches 349; Indels 0; Gaps 0;
Matches 42; Conservative 157;

QY 130 TCTATCCAGTTTACTGTGTAACGCTCTAATCCAAATCGCCAGATTGTGTAATGG 189
DB 554 YMW..K.B.G.DGW.SR.HEKMB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATTGCTGGCGGATAGTGTATGACGAAGATATGACCGCTCATGTGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBA.RT.NRNG.C.DBAGH.AC.W..CR..NSTY.....TYG.. 435
QY 250 GATGAGGATTTTGTATGACGAGCAGCAGTATTGTGCGCAAAAGTCCATCTTATCT 309
DB 434 GW.T..HB.W..MENT..N..C.AABDSNBS.MT.T..HR..MDCD..R.MAC.CM.BT 375
QY 310 AATTATACGCGATCCGCTGTGTTATGACGACCCCAAGCGCATCCGATGTTGAGT 369
DB 374 ..S..BR.NSG.H.TSSGH.YCB..DHN..AM.WM..HNSN.NMC.D....WYDDT.Y 315
QY 370 GTCAGTATTTCCATTCAGCTTTATTAAACCGTGAATGTCAAACAATAGAAATTT 429
DB 314 H..ANTS...H..ND.B..R..MTH.SN..KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGTATGACCCCGGAGCGGCTAATGTTGATCTTCACTTACTTACCAAGAAGAAATG 489
DB 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDMB.TKCSY....S.RSNECMK 195
QY 490 CCGGTGCTTTTGTATGATGTCGCGCAATTCAGCATCGTGGCAAAATTAACGCGTGG 549
DB 194 YG.KTHGCBTC.Y..CHB.MT.HYWA.T.HDSDSC.SR.NB.W..CSNBTH..YY. 135
QY 550 GTGAATACACATCCGCTGCTACTATTCTTGAAGAAAAGATTTTATGCGCGATATT 609
DB 134 A...HHHR.C.HRYRNMWS.R.CC.NH..CM.CRHRMNY.H.HM.B.RTYN.M..HC 75
QY 610 ATTGATGCTTACGCGCAATTTGGCGTGAAGTTATGCGATGCACTTAAACGCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSB.C..SR.H.HACB..TKRB...G...YM 15
QY 670 ATCAATAC 677
DB 14 .HY.RHHM 7

RESULT 6
US-10-123-155-426/C
; Sequence 426, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Deenoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary B.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven

;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P330R1C30
;; CURRENT APPLICATION NUMBER: US/10/123,155
;; CURRENT FILING DATE: 2002-04-15
;; Prior Application removed - See Palm or File Wrapper
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 426
;; LENGTH: 747
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-123-155-426

Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5; Mismatches 349; Indels 0; Gaps 0;
Matches 42; Conservative 157;

QY 130 TCTATCCAGTTTACTGTGTAACGCTCTAATCCAAATCGCCAGATTGTGTAATGG 189
DB 554 YMW..K.B.G.DGW.SR.HEKMB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATTGCTGGCGGATAGTGTATGACGAAGATATGACCGCTCATGTGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBA.RT.NRNG.C.DBAGH.AC.W..CR..NSTY.....TYG.. 435
QY 250 GATGAGGATTTTGTATGACGAGCAGCAGTATTGTGCGCAAAAGTCCATCTTATCT 309
DB 434 GW.T..HB.W..MENT..N..C.AABDSNBS.MT.T..HR..MDCD..R.MAC.CM.BT 375
QY 310 AATTATACGCGATCCGCTGTGTTATGACGACCCCAAGCGCATCCGATGTTGAGT 369
DB 374 ..S..BR.NSG.H.TSSGH.YCB..DHN..AM.WM..HNSN.NMC.D....WYDDT.Y 315
QY 370 GTCAGTATTTCCATTCAGCTTTATTAAACCGTGAATGTCAAACAATAGAAATTT 429
DB 314 H..ANTS...H..ND.B..R..MTH.SN..KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGTATGACCCCGGAGCGGCTAATGTTGATCTTCACTTACTTACCAAGAAGAAATG 489
DB 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDMB.TKCSY....S.RSNECMK 195
QY 490 CCGGTGCTTTTGTATGATGTCGCGCAATTCAGCATCGTGGCAAAATTAACGCGTGG 549
DB 194 YG.KTHGCBTC.Y..CHB.MT.HYWA.T.HDSDSC.SR.NB.W..CSNBTH..YY. 135
QY 550 GTGAATACACATCCGCTGCTACTATTCTTGAAGAAAAGATTTTATGCGCGATATT 609
DB 134 A...HHHR.C.HRYRNMWS.R.CC.NH..CM.CRHRMNY.H.HM.B.RTYN.M..HC 75
QY 610 ATTGATGCTTACGCGCAATTTGGCGTGAAGTTATGCGATGCACTTAAACGCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSB.C..SR.H.HACB..TKRB...G...YM 15
QY 670 ATCAATAC 677
DB 14 .HY.RHHM 7

RESULT 7
US-10-146-731-426/C
; Sequence 426, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Deenoyers, Luc

```
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Mei-Qiang
APPLICANT: Gerltisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C323
CURRENT APPLICATION NUMBER: US/10/146,731
PRIOR APPLICATION removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-146-731-426
```

```
Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;
```

```
QY 130 TCTATCCAGGTTTACTGCTGTAACCGCTTAATCCAAATCGCCAGATTGGTAATGG 189
DB 554 YMW..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATTGCGTGGGGAGTAGTATGACGAAGATAGACCGCTCAGTGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBAI.RT.NRKNG.C.DBAQH.AC.M..CR..NSTYY.....TYG.. 435
QY 250 GATGAGATTGTTGATGACGAGACGACGATTTGTGCGCAAAAGTCATTAATCTT 309
DB 434 GW.T.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MWCDD..R.MAC.CM.BT 375
QY 310 AATTATGACGATCCGCTGCTGTTGATGCAACCCCAAGCGATCCGAATGTTGAGT 369
DB 374 .S...BR.NSG.H.TSSGH.YCB..DHNM..AM.MM..HNSN.NMC.D...WYDDT.Y 315
QY 370 GTGAGATTTCCTTATTAACCGCTGATGTCACAAATAGAGATTTT 429
DB 314 H..ANTS...H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGATGACCCCGAGCGCGCTAATGTTGATCTTCACTTAATCTCAAGAGAATG 489
DB 254 T.TA..SNB.NCAM..ABASYBD.KT.SB.S.MSYHCDRMB.TKCSY....S.RSNBCKM 195
QY 490 CCGCTGCTTTGATGATGTCGCGCAATTCAGATGCTGGCAAAATTAAGCGCTGCG 549
DB 194 YG.KYHHCGBTC.Y..CHB.MT.HYWA.T.HGDSGC.SR.NB.W..CSNBTH..Y.Y. 135
QY 550 GTTGAATACACATCCGCTGATCTATTCTTATAGAAAAGATTTTATGTCGGATAT 609
DB 134 .A...HHHR.C.HRYRBNWS.R.CC.NH..CM.CMEHMMY.H.HM.B.RTYN.M..HC 75
QY 610 ATTGATGCTAGCCCAATTTGGCGCGCAAGTTATCGATGACCATTAAGCCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTPMSBC..SR.H.HACB..TKRB...G...YM 15
QY 670 ATCAATAC 677
DB 14 .HY.RHMM 7
```

RESULT 8
US-10-140-472-426/c

```
Sequence 426, Application US/10140472
Publication No. US20030138888A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Mei-Qiang
APPLICANT: Gerltisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C168
CURRENT APPLICATION NUMBER: US/10/140,472
PRIOR APPLICATION removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-472-426
```

```
Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;
```

```
QY 130 TCTATCCAGGTTTACTGCTGTAACCGCTTAATCCAAATCGCCAGATTGGTAATGG 189
DB 554 YMW..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATTGCGTGGGGAGTAGTATGACGAAGATAGACCGCTCAGTGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBAI.RT.NRKNG.C.DBAQH.AC.M..CR..NSTYY.....TYG.. 435
QY 250 GATGAGATTGTTGATGACGAGACGACGATTTGTGCGCAAAAGTCATTAATCTT 309
DB 434 GW.T.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MWCDD..R.MAC.CM.BT 375
QY 310 AATTATGACGATCCGCTGCTGTTGATGCAACCCCAAGCGATCCGAATGTTGAGT 369
DB 374 .S...BR.NSG.H.TSSGH.YCB..DHNM..AM.MM..HNSN.NMC.D...WYDDT.Y 315
QY 370 GTGAGATTTCCTTATTAACCGCTGATGTCACAAATAGAGATTTT 429
DB 314 H..ANTS...H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGATGACCCCGAGCGCGCTAATGTTGATCTTCACTTAATCTCAAGAGAATG 489
DB 254 T.TA..SNB.NCAM..ABASYBD.KT.SB.S.MSYHCDRMB.TKCSY....S.RSNBCKM 195
QY 490 CCGCTGCTTTGATGATGTCGCGCAATTCAGATGCTGGCAAAATTAAGCGCTGCG 549
DB 194 YG.KYHHCGBTC.Y..CHB.MT.HYWA.T.HGDSGC.SR.NB.W..CSNBTH..Y.Y. 135
QY 550 GTTGAATACACATCCGCTGATCTATTCTTATAGAAAAGATTTTATGTCGGATAT 609
DB 134 .A...HHHR.C.HRYRBNWS.R.CC.NH..CM.CMEHMMY.H.HM.B.RTYN.M..HC 75
QY 610 ATTGATGCTAGCCCAATTTGGCGCGCAAGTTATCGATGACCATTAAGCCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTPMSBC..SR.H.HACB..TKRB...G...YM 15
```


QY 670 ATCATAC 677
DB 14 .HY.RHNM 7

RESULT 9

US-10-141-761-426/c
Sequence 426, Application US/10141761
Publication No. US20030148432A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C198
CURRENT APPLICATION NUMBER: US/10/141,761
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 426
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-141-761-426

Query Match

4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCATCCAGGTTTACTGTGAAACGCTTAATCCAAATGCGCAGATTGCTAATG 189
DB 554 YMW..K.B.G.DGW.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATGCGTGGCGGATAGTATGACGAAGATATGACCGCTCATGCTGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBAH.RT.NRNG.C.DBAGH.AC.M..CR..NSTY.....TYG.. 435
QY 250 GATGAGATTGTTGATGACGACGACGATATTGTGCGCAAAAGTCATCTTACTCT 309
DB 434 GW.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MDCD..R.MAC.CM.BT 375
QY 310 AATTTCACGCGATCCGCTGTTGATGCGAACCCCAAGCGCATCGGATGTTGAGT 369
DB 374 .S...BR.NSG.H.TSSGH.YCB..DHNM..AM.MM...HNSN.NMC.D....WYDT.Y 315
QY 370 GTCGATTTCCATTCAGCTTATTAACCCGTGATGTCAAAAGTAAAGATTTT 429
DB 314 H..ANTS....H.ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGTATGACCCCGGCGCGCTAATTGTTGATCTTCACTTACTCAAGAGAAATG 489
DB 254 T.TA..SNB.NCAM..ABASYB.XT.SB.S.MSYHCDRMB.TKCSY...S.RSNBCKK 195
QY 490 CCGTGCTTTGATGATCGCGCAATTCAGATGCGGCAAAATTAACGCGTGG 549
DB 194 YG.KYHCGMBTC.Y..CHB.MT.HYVA.T.HDSDSC.SR.NB.W..CSNBTH..YY. 135
QY 550 GTTGAATACACATCGGTGATCTATTTTCATTAGAAAAAGTTTAACTGCGCATATT 609

DB 134 A...HHHR.C.HRYRBNWS.R.CC.NH..CM.CENHRMRY.H.HM.B.RTN.M..HC 75
QY 610 ATTATGCTTACCGCAATTTGGCGTCAAGTTAATGCAATGACCAATTAACCCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTMSBC..SR.H.HACB..TKRB...G...YM 15
QY 670 ATCATAC 677
DB 14 .HY.RHNM 7

RESULT 10

US-10-142-885-426/c
Sequence 426, Application US/10142885
Publication No. US20030157604A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C248
CURRENT APPLICATION NUMBER: US/10/142,885
CURRENT FILING DATE: 2002-05-10
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 426
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-142-885-426

Query Match

4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCATCCAGGTTTACTGTGAAACGCTTAATCCAAATGCGCAGATTGCTAATG 189
DB 554 YMW..K.B.G.DGW.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATGCGTGGCGGATAGTATGACGAAGATATGACCGCTCATGCTGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBAH.RT.NRNG.C.DBAGH.AC.M..CR..NSTY.....TYG.. 435
QY 250 GATGAGATTGTTGATGACGACGACGATATTGTGCGCAAAAGTCATCTTACTCT 309
DB 434 GW.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MDCD..R.MAC.CM.BT 375
QY 310 AATTTCACGCGATCCGCTGTTGATGCGAACCCCAAGCGCATCGGATGTTGAGT 369
DB 374 .S...BR.NSG.H.TSSGH.YCB..DHNM..AM.MM...HNSN.NMC.D....WYDT.Y 315
QY 370 GTCGATTTCCATTCAGCTTATTAACCCGTGATGTCAAAAGTAAAGATTTT 429
DB 314 H..ANTS....H.ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGTATGACCCCGGCGCGCTAATTGTTGATCTTCACTTACTCAAGAGAAATG 489

```
Db 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDRMB.TKCSY....S.RSBNCKM 195
Qy 490 CCGCTGCTTTTATCATGTGCGCAATTCAGATGCGTGGCAAAATTAAGCGCTGG 549
Db 194 YG.KYHHCNBT.C.Y..CHB.MT.HYWA.T.HGDSGC.SR.NB.W..CSNBTH..YY. 135
Qy 550 GTTGAATACACATCCGTGTACTATTTCATTAAGAAAAGATTTTATGTCGGATATT 609
Db 134 A....HHR.C.HRYRWRWS.R.CC.NH..CM.CMHRMY.H.HM.B.RTYN.M..HC 75
Qy 610 ATTGATGCCAGCCCAATTTGGCGTCAAGTTAATCCATACCATTAAGCGCTGG 669
Db 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSB.C..SR.H.HACB..TKRB...G...YM 15
Qy 670 ATCAATAC 677
Db 14 .HY.RHMM 7

RESULT 11
US-10-158-790-426/c
; Sequence 426, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; CURRENT FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 426
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-426

Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

Qy 130 TCTATCCAGTTTACTGTGTAAGCGCTTAATCCAAATCGGCCAGATTGTGTAATGG 189
Db 554 YWM..K.B.G.DGW.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
Qy 190 GCATTGCGTGGCGGATAGTATGACGAAGATATGACCGCTCATGTGAGAACTGTC 249
Db 494 DC.R.D.NS.N.BBA..RT..NRKNG.C.DBAGH.AC.M..CR..NSTY.....TYG.. 435
Qy 250 GATGAGATTTTATGATGACGAGCAGATATTTGTGCGCAAAAGTCCATATTATCT 309
Db 434 GW.T..HB.W..MHT..N..C.AABDNBS.MT.T..HR..MMCD..R.MAC.CM.BT 375
Qy 310 AATTATCAGGACCGGTGTGTATGATGACCAACCCCAAGCATCCGATGTTGAGT 369
Db 374 S...BR.NSG.H.TSGH.YCB..DHMM..AM.MM..HNSN.NMC.D....WYDPT.Y 315
```

```
Qy 370 GTGAGATTTTCCATTAACGGTTTATTAACCGGTGGAATGCAACAATAGAAGATT 429
Db 314 H..ANTS...H..ND.B..R..MTH..SN...KOS.Y.YA.SA.GHCCMG.A..WC.CT 255
Qy 430 GGTATGACCCCGAGCGCGGTAAATTTGATCTTCACTTACTTACTCAAGAAGATG 489
Db 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDRMB.TKCSY....S.RSBNCKM 195
Qy 490 CCGCTGCTTTTATCATGTGCGCAATTCAGATGCGTGGCAAAATTAAGCGCTGG 549
Db 194 YG.KYHHCNBT.C.Y..CHB.MT.HYWA.T.HGDSGC.SR.NB.W..CSNBTH..YY. 135
Qy 550 GTTGAATACACATCCGTGTACTATTTCATTAAGAAAAGATTTTATGTCGGATATT 609
Db 134 A....HHR.C.HRYRWRWS.R.CC.NH..CM.CMHRMY.H.HM.B.RTYN.M..HC 75
Qy 610 ATTGATGCCAGCCCAATTTGGCGTCAAGTTAATCCATACCATTAAGCGCTGG 669
Db 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSB.C..SR.H.HACB..TKRB...G...YM 15
Qy 670 ATCAATAC 677
Db 14 .HY.RHMM 7

RESULT 12
US-10-137-871-426/c
; Sequence 426, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; SEQ ID NO 426
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-871-426

Query Match 4.4%; Score 35; DB 16; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

Qy 130 TCTATCCAGTTTACTGTGTAAGCGCTTAATCCAAATCGGCCAGATTGTGTAATGG 189
Db 554 YWM..K.B.G.DGW.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
Qy 190 GCATTGCGTGGCGGATAGTATGACGAAGATATGACCGCTCATGTGAGAACTGTC 249
Db 494 DC.R.D.NS.N.BBA..RT..NRKNG.C.DBAGH.AC.M..CR..NSTY.....TYG.. 435
```

QY 250 GATGAGGATTTTGTGAGGAGGAGACGACGATTTTGGGCAAAAAGGTCGACTTATCCT 309
 Db 434 G M T . H B W . M E N T . N . C . A B D S N B S . M T T . H R . M M C D D . R . M A C . C M . B T 375
 QY 310 AATTTATCAGCAGTCGCGCTGGTGTGAGCAACCCCAACGCGATCGGATGGTTGAGT 368
 Db 374 . S . . . B R . N S G . H . T S S G H . Y C B . D H N M . A M . N M . . . H N S N . N M C . D . . . W Y D P T . Y 315
 QY 370 GTGAGTATTTCCGATTAACGCTTTATTTAAACCCGCGTGAATGTCCAAACAAATAGAGATTT 429
 Db 314 H . A N T S . . . H . N D . B . R . M T H . S N . . K C S . Y . Y A . S A . G H C M S G . A . W C . C T 255
 QY 430 GGTATCGACCCCGGAGCGCGCTAATTTGGTTGATCTTCATCTTACTTCACAAAGAGAAATG 489
 Db 254 T . T A . . S N B . N C A M . A B A S Y B . K T . S B . S . M S Y H O R M B . T K S Y . . . S . R S N B O M K 195
 QY 490 CCGCTGCGCTTTTGATCATGTGCGGCAATTCAGCATGCGTGCGCAAAAATTCGCGCGCTCGC 549
 Db 194 Y G . K Y H N C G N B T C . Y . C H B . M T . H Y M A . T . H D G S G C . S R . N B . W . C S N B T H . . Y Y 135
 QY 550 GTTGATACACATCCGCGTGACTATTTTCATTAGAAAAAGATTTTGTAGCGCGATATT 609
 Db 134 . A . . . H H N R . C . H R R R M W S . R . C C . N H . C M . C R M R M Y . H . H M . B . R T Y N . M . H C 75
 QY 610 ATTGATCCCTACGCCCAATTTGGCGTGCAGTTAATTCGATGACCATTCATTAACGCGCGCTTG 668
 Db 74 D T N . N N . N M H . D . M . T . M M . T . Y . Y . S T A M S B C . S R . H . H A C B . T K R B . . G . . . Y M 15
 QY 670 ATCAATAC 677
 Db 14 . H Y . R H E M 7

```

RESULT 13
US-10-140-923-426/c
; Sequence 426, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:

```

```

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zenlin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P333081C188
CURRENT APPLICATION NUMBER: US/10/140,923
CURRENT FILING DATE: 2002-05-07
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 426
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-923-426

```

Query Match 4.4%; Score 35; DB 16; Length 747;
 Best Local Similarity 7.7%; Pred. No. 5.5;
 Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0

130 TCTATCAGGTTTACTGCTGAAGGCTCTATCTCAATGCGCGAGATTGCTAAATCG 189

```

Db      554 YMW. K. B. G. DGM. SR. HBKXB. NDMY. AB. N. AA. A. SBBS. NK. H. .... TSD 499
Oy      190 GCAITGCGTGGCGGGATAGTGTATGATACGAAGATATGACCGCTCATGTGTGGAACTGTG 248
Db      494 DC. R. D. NS. N. BBAA. RT. N. RKNG. C. DBAGH. AC. M. CR. NSTYV. .... TYG. 435
Oy      250 GATGAGGATTTTGATGACGACGAGACGAGCGATTTGTGGGCAAAAAGTCATATTATTCCT 309
Db      434 GW. T. HB. W. MENT. N. C. AABDSNB. MT. T. HR. VMCD. R. MAC. CM. BT 375
Oy      310 AATTTATACGCGATCCGCTGTTGATGCAACCCCAACCGATCCGATGTTGAGT 365
Db      374 S. . BR. NSG. H. TSGH. YCB. . DHNV. AM. NW. . HNSN. NMC. D. . . WYDPT. Y 318
Oy      370 GTGAGATTTCCCATTCAGCTTATTTAAACCCGCTGGAATGTCAACAAATGAGATTTT 422
Db      314 H. ANT. S. . H. ND. B. R. MTH. SN. . KCS. Y. YA. SA. GHCMG. A. WC. CT 255
Oy      430 GGTATGACACCCCGACGCGCTAATGATGTTGATCTTACTTACTTACTCAAGAAGAAATG 488
Db      254 T. TH. . SNB. NCAM. ABASYBD. KT. SB. S. MSYHCBMB. TKSY. . . S. RSNBCK 195
Oy      490 CCGCTGCGTTTGATATGATGTCCGCAAAATTCAGCATGCTGGCAAAAATTCGCGGCTG 544
Db      194 YG. KYHNCGBTC. Y. CHB. MT. HYWA. T. HDGSGC. SR. NB. W. CSNBTH. YY. 135
Oy      550 GTTGAAATACACATCCGCTGTACTATTTCATTTATGAAAAAGTTTGTAGCGCGATAT 608
Db      134 A. . . . HHR. C. HRRRWMS. R. CC. NH. . CM. CRHRMY. H. HM. B. RTYN. M. HC 75
Oy      610 ATGTAGGCTTACGCCAAATTTGGCGTCAAGTTAATGCATGACCATTTAAAGCGCGCTG 666
Db      74 DTN. NN. NWH. D. M. T. MM. T. Y. YSSTAMSBC. SR. H. HACB. TKRB. . G. . YM 15
Oy      670 ATGCAATAC 677
Db      14 . HY. RHNM 7

```

RESULT 14
US-10-141-756-426/c
; Sequence 426, Application US/10141756
; Publication No. US20030207359A1

```

1 GENERAL INFORMATION:
2 APPLICANT: Baker, Kevin P.
3 APPLICANT: Beresini, Maureen
4 APPLICANT: Desrogers, Laura
5 APPLICANT: Desnoyers, Luc
6 APPLICANT: Filvaroff, Ellen
7 APPLICANT: Gao, Wei-Qiang
8 APPLICANT: Gerritsen, Mary E.
9 APPLICANT: Goddard, Audrey
10 APPLICANT: Godowski, Paul J.
11 APPLICANT: Gunney, Austin L.
12 APPLICANT: Sherwood, Steven
13 APPLICANT: Smith, Victoria
14 APPLICANT: Stewart, Timothy A.
15 APPLICANT: Tumas, Daniel
16 APPLICANT: Watanabe, Colin K
17 APPLICANT: Wood, William
18 APPLICANT: Zhang, Zemin
19 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
20 TITLE OF INVENTION: ACIDS ENCODING THE SAME
21 FILE REFERENCE: P3330R1C200
22 CURRENT APPLICATION NUMBER: US/10/141,756
23 PRIORITY APPLICATION removed - 2002-05-08
24 NUMBER OF SEQ ID NOS: 550
25 SEQ ID NO 426
26 LENGTH: 747
27 TYPE: PRT
28 ORGANISM: Homo Sapien

```

US-10-141-756-426

Query Match 4.4%; Score 35; DB 16; Length 747;

Best Local Similarity 7.7%; Pred. No. 5.5; Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCTATCCAGGTTTTCGTGGTGAACGCTCTAATCCAGTCCGACAGATGGTGAATG 189
 DB 554 YWM..K.B.G.DGW..SR.HBKWB.NDMY..AB..N.AA.A.SBSB.NK.H.....TSD 495
 QY 190 GCATTGCGCTGGCGGATAGTGTATGACGAGATATGACCGCTCATGTGGAGAACCTGTC 249
 DB 494 DC.R.D..NS.N.BBA..RT..NRKNG..C.DBAGH..AC.M..CR..NSTYV.....TYG.. 435
 QY 250 GATGAGATTGATGACGAGACGATATTGTCGCAAAAAGTCCATATCTATCT 309
 DB 434 GW.T..HB.W..MENT..N..C.AABDSNBS..MT.T..HR..WMCDD..R..MAC..CM..BT 375
 QY 310 AATTATACGCGATCCGCTGTGTATGACCAACCCCAACGCGATCCGATGCTGAGT 369
 DB 374 ..S..BR..NSG..H.TSSGH..YCB..DHNW..AM..NMW..HNSN..NMC.D....WYDPT..Y 315
 QY 370 GTGAGATTTCCTATGACGCTTATTAACCGGTGATGTCACAAAGATGATTT 429
 DB 314 H..AMTS...H..ND.B..R..MTH..SN..KCS..Y..YA..SA..GHCMG..A..WC..CT 255
 QY 430 GGTATGACCCCGACGCGCTAATGTTGATCTTCACTTACTTCAAGAAGAATG 489
 DB 254 T..TA..SNB..NCAM..ABASVBD..KT..SB.S..MSYHCDMB..TKCSY....S..RSNBCKM 195
 QY 490 CCGCTGGCTTTGATCATGTCCGCAATTCAGCATGCGTGCAAAAATTCGCGCTGCG 549
 DB 194 YG..KTHCGNBT..Y..CHB..MT..HYWA..T..HDGSGC..SR..NB..W..CSNBTH..YY.. 135
 QY 550 GTTGAATACATCCGCTGTGATCTTATTCATTAAGAAAAGTTTGTAGCGGATATT 609
 DB 134 A...A...HHNR..C.HRRRWMS..R..CC..NH..CM..CRHMMY..H..HM..B..RTYN..M..HC 75
 QY 610 ATTGATGCTTACGCAAAATTTGGCGTGAAGTTAATGATGACCATTAAGCGCGCTG 669
 DB 74 DTN..NN..NMH..D.M.T..MM..T..Y..YSSYMSBC..SR..H..HACB...TKRB...G...YM 15
 QY 670 ATCAATAC 677
 DB 14 ..HY..RHHM 7
 RESULT 15
 US-10-141-756-426/c
 Sequence 426, Application US/10141759
 Publication No. US20030207361A1
 GENERAL INFORMATION:
 APPLICANT: Baker, Kevin P.
 APPLICANT: Beresini, Maureen
 APPLICANT: DeForge, Laura
 APPLICANT: Desnoyers, Luc
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Sherwood, Steven
 APPLICANT: Smith, Victoria
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Watanabe, Colin K
 APPLICANT: Wood, William
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 TITLE OF INVENTION: ACIDS ENCODING THE SAME
 FILE REFERENCE: P3330RIC197
 CURRENT APPLICATION NUMBER: US/10/141,759

CURRENT FILING DATE: 2002-05-08

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 426

LENGTH: 747

TYPE: PRT

ORGANISM: Homo Sapien

US-10-141-756-426

Query Match 4.4%; Score 35; DB 16; Length 747;

Best Local Similarity 7.7%; Pred. No. 5.5; Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCTATCCAGGTTTTCGTGGTGAACGCTCTAATCCAGTCCGACAGATGGTGAATG 189
 DB 554 YWM..K.B.G.DGW..SR.HBKWB.NDMY..AB..N.AA.A.SBSB.NK.H.....TSD 495
 QY 190 GCATTGCGCTGGCGGATAGTGTATGACGAGATATGACCGCTCATGTGGAGAACCTGTC 249
 DB 494 DC.R.D..NS.N.BBA..RT..NRKNG..C.DBAGH..AC.M..CR..NSTYV.....TYG.. 435
 QY 250 GATGAGATTGATGACGAGACGATATTGTCGCAAAAAGTCCATATCTTATCT 309
 DB 434 GW.T..HB.W..MENT..N..C.AABDSNBS..MT.T..HR..WMCDD..R..MAC..CM..BT 375
 QY 310 AATTATACGCGATCCGCTGTGTATGACCAACCCCAACGCGATCCGATGCTGAGT 369
 DB 374 ..S..BR..NSG..H.TSSGH..YCB..DHNW..AM..NMW..HNSN..NMC.D....WYDPT..Y 315
 QY 370 GTGAGATTTCCTATGACGCTTATTAACCGGTGATGTCACAAAGATGATTT 429
 DB 314 H..AMTS...H..ND.B..R..MTH..SN..KCS..Y..YA..SA..GHCMG..A..WC..CT 255
 QY 430 GGTATGACCCCGACGCGCTAATGTTGATCTTCACTTACTTCAAGAAGAATG 489
 DB 254 T..TA..SNB..NCAM..ABASVBD..KT..SB.S..MSYHCDMB..TKSY....S..RSNBCKM 195
 QY 490 CCGCTGGCTTTGATCATGTCCGCAATTCAGCATGCGTGCAAAAATTCGCGCTGCG 549
 DB 194 YG..KTHCGNBT..Y..CHB..MT..HYWA..T..HDGSGC..SR..NB..W..CSNBTH..YY.. 135
 QY 550 GTTGAATACATCCGCTGTGATCTTATTCATTAAGAAAAGTTTGTAGCGGATATT 609
 DB 134 A...A...HHNR..C.HRRRWMS..R..CC..NH..CM..CRHMMY..H..HM..B..RTYN..M..HC 75
 QY 610 ATTGATGCTTACGCAAAATTTGGCGTGAAGTTAATGATGACCATTAAGCGCGCTG 669
 DB 74 DTN..NN..NMH..D.M.T..MM..T..Y..YSSYMSBC..SR..H..HACB...TKRB...G...YM 15
 QY 670 ATCAATAC 677
 DB 14 ..HY..RHHM 7
 Search completed: April 30, 2004, 23:19:44
 Job time : 369.368 Secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 08:18:22 ; Search time 2238.64 Seconds

(without alignments)
10745.552 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagatattcttcgttca.....cttcacagagatgctgtc 555

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_da.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vt.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vt.*
30: em_hg_hum.*
31: em_hg_inv.*
32: em_hg_other.*
33: em_hg_mus.*
34: em_hg_pln.*
35: em_hg_rtd.*
36: em_hg_mam.*
37: em_hg_vtc.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 213.8 | 38.5 | 247450 | 1 | AP005349 |
| 2 | 213.8 | 38.5 | 302331 | 1 | AB016810 |
| 3 | 191.4 | 34.5 | 10977 | 1 | AB004352 |
| 4 | 178.8 | 32.2 | 339350 | 1 | AB005088 |
| 5 | 64.8 | 11.7 | 622 | 1 | VC0231080 |
| 6 | 37.4 | 6.7 | 178879 | 2 | AC124052 |
| 7 | 37.2 | 6.7 | 2000 | 6 | AX655393 |
| 8 | 36.8 | 6.6 | 26997 | 6 | AX251505 |
| 9 | 36.8 | 6.6 | 183397 | 10 | AL592225 |
| 10 | 36.4 | 6.6 | 209061 | 2 | AC130676 |
| 11 | 36.2 | 6.5 | 54762 | 2 | AC130294 |
| 12 | 36 | 6.5 | 101342 | 8 | AP003850 |
| 13 | 36 | 6.5 | 271370 | 2 | AC112046 |
| 14 | 35.8 | 6.5 | 27243 | 3 | CE006547 |
| 15 | 35.8 | 6.5 | 215278 | 10 | AC0659562 |
| 16 | 35.6 | 6.4 | 115218 | 9 | HS15941 |
| 17 | 35.6 | 6.4 | 170260 | 2 | AL158204 |
| 18 | 35.6 | 6.4 | 212449 | 2 | AL355972 |
| 19 | 35.6 | 6.4 | 224330 | 2 | AC113079 |
| 20 | 35.2 | 6.3 | 671 | 8 | AY135373 |
| 21 | 35.2 | 6.3 | 169572 | 2 | AC023832 |
| 22 | 35 | 6.3 | 128769 | 2 | AL139261 |
| 23 | 35 | 6.3 | 170110 | 9 | CNS01RHT |
| 24 | 35 | 6.3 | 215747 | 10 | AC117581 |
| 25 | 34.8 | 6.3 | 623 | 3 | AY047332 |
| 26 | 34.6 | 6.2 | 7218 | 6 | 166494 |
| 27 | 34.6 | 6.2 | 103513 | 9 | AL357713 |
| 28 | 34.6 | 6.2 | 198105 | 2 | AC026160 |
| 29 | 34.6 | 6.2 | 233018 | 2 | AC113788 |
| 30 | 34.6 | 6.2 | 237754 | 2 | AC117125 |
| 31 | 34.6 | 6.2 | 288092 | 2 | AC118793 |
| 32 | 34.4 | 6.2 | 2277 | 6 | AR004980 |
| 33 | 34.4 | 6.2 | 244169 | 2 | AC107357 |
| 34 | 34.4 | 6.2 | 245308 | 2 | AC108307 |
| 35 | 34.4 | 6.2 | 262126 | 2 | AC098502 |
| 36 | 34.2 | 6.2 | 3596 | 8 | AF096277 |
| 37 | 34 | 6.1 | 131027 | 9 | AC004806 |
| 38 | 33.8 | 6.1 | 236470 | 10 | AL831793 |
| 39 | 33.8 | 6.1 | 51531 | 2 | AC115089 |
| 40 | 33.8 | 6.1 | 143589 | 10 | AC124745 |
| 41 | 33.8 | 6.1 | 174508 | 9 | AC021561 |
| 42 | 33.8 | 6.1 | 177781 | 9 | AC114689 |
| 43 | 33.8 | 6.1 | 268699 | 2 | AC139758 |
| 44 | 33.4 | 6.0 | 1629 | 8 | FSY420192 |
| 45 | 33.4 | 6.0 | 55096 | 2 | AC100894 |

ALIGNMENTS

RESULT 1
LOCUS AP005349 247450 bp DNA linear BCT 04-DEC-2003
DEFINITION Vibrio vulnificus YJ016 DNA, chromosome II, complete genome,
section 6/8.
ACCESSION AP005349 BA000038
VERSION AP005349.1 GI:37201329
KEYWORDS
SOURCE Vibrio vulnificus YJ016
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
1 Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L.,
Lin, Y.M., Chen, H.J., Shen, A.B., Li, D.C., Su, T.D., Shao, C.P.,

TITLE Lee, C.T., Hor, L.I. and Tsai, S.F.
JOURNAL Comparative Genome Analysis of *Vibrio vulnificus*, a Marine Pathogen
REFERENCE Genome Res. 13, 2577-2587 (2003)
AUTHORS 2 (bases 1 to 247450)
JOURNAL Chen, C.Y., Wu, K.M. and Tsai, S.F.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Shih Feng Tsai, National Health Research
Institutes, Division of Molecular and Genomic Medicine, 128
Yen-Chiu-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China
(E-mail: pctsa@nhri.org.tw, Tel: 886-2-8146-1041,
Fax: 886-2-2789-0484)
COMMENT This sequence was determined by the Sequencing Core of the National
Yang-Ming University Genome Research Center (YMC).
http://genome.ym.edu.tw

FEATURES
source
location/Qualifiers
1..247450
/organism="Vibrio vulnificus Y0016"
/mol_type="genomic DNA"
/strain="Y0016"
/db_xref="taxon:196600"
/chromosome="11"
complement(73..1092)
/gene="VVA1126"
complement(73..1092)
/gene="VVA1126"
codon_start=1
/transl_table=11
/product="outer membrane protein N, non-specific porin"
/protein_id="BAC97152.1"
/db_xref="GI:37201330"
/translation="MNKIIALAVASATFGTQAVAEVYNDGTTFSIGHVSVALD
SKQDLGVAVSPRINFNATIDLGNGFPADAGKEMSYLDGGENALSTRIGYGLTH
ESLGRGVGTQAWPYVAVGADMPAFANDPFIYODHNLGTGRAEEMVSYANFELG
BAGKALIGLGMGRKTDNNNSYGNRAQIALNYSITANYANTAYGDNVYGVGKGT
SSSHTSATSGYSGKGLIACVYANETNMSGDGLDPTSAERALAATYTNKINIS
VYVESVIDEELSKRYSTSAVAETNFTPKVGFAGYFDLQSGSDIKKADKRMGLG
ARYYL"
complement(1331..1519)
/gene="VVA1127"
complement(1331..1519)
/gene="VVA1127"
/note="identified by Glimmer2"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC97153.1"
/db_xref="GI:37201331"
/translation="MRALYVMDISLKLASRGYLGIRPAFNPFSLIKGESISVCNII
INISVFLITPNVYVCF"
1559..2050
/gene="VVA1128"
1559..2050
/gene="VVA1128"
codon_start=1
/transl_table=11
/product="thiol-disulfide isomerase and thioredoxins"
/protein_id="BAC97154.1"
/db_xref="GI:37201332"
/translation="MKIVILKEMLPVLAIVISALDFYSSNLPKRNALPLSAMT
TLGTVDLSAQSGSPVAVYFWATWCGPCKVSPPTINIDSPYISVAMSGDEGSV
NOYLOSQYVDYDITNDPTGKISRDMGVQVTPVIVIRNGKIEFTISGVSPDGLMLR
LWS"
2203..3180
/gene="VVA1129"
2203..3180
/gene="VVA1129"
codon_start=1
/transl_table=11
/product="putative outer membrane protein"
/protein_id="BAC97155.1"
/db_xref="GI:37201333"
/translation="MKKALATPAIISAVVTGSSPPATYTKTDGTELUKIGAVFERGDP
IGTDKVEIDGTMEDSTRALINVKGTITDLGNDLQAFGYEARQKTGESSFNRRYVAG

VTNAGASVGVKQDNASVLSLTDITERSGVQOPIDASDSDTSTPRVRSFDALQI
EATYSASDNDNDLNGISGLYSLPIGLDGLAVSTGDNAGNGSKQILVAGYVND
LYLGATFSTGDDDDAKQEFYAEFVQYKISKPSVAMYPAEDDNGTSDATKG
LELVGYKLNNSFRITLYSLNKLDDVKNKSVTSGEDTLRGVRYDF"
3606..5381
/gene="VVA1130"
3606..5381
/gene="VVA1130"
/note="identified by GeneMark and Glimmer2"
codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC97156.1"
/db_xref="GI:37201334"
/translation="MYDGRSKIVANSKYLIIKAYIVAVLITQTLVDERTLPSY
VPAIMVALVSPSHKIKVICALIATSTLFGGVDDPIDIVETITFLILPCYLL
PFGTIAVFAVALLTASFTYVPPGDLBILDALELVISAFAISIMVYQKFIYQL
TKYREDSTDTYTLQENRAFYSDINQVBSLAQQQRNCALNIDIDAFKEINSEFGE
QADLILKRSRRLRKNTLGFQGVYMEGDEPVALYVEKSIDEDVALDKIKRALTA
APFLERSYQISVATGIALFNESMTNLDLNSDIAQAKATRGSGVAVRDEALD
ETIRQIQRERDKATELKQPSLVQPKVITITGHTGAEALIRMOHEPILGITPEP
ITIAEKTKOITIGRWITLBAQQAQKMLDLYKIPVSNVSTQPGAHDDITLOPETA
LASHLPASLTQDEITETVLMKPKKLEITGNLRCBGTVVAIDPGIAYSSLNTRKO
LPIDVAKIDKAPIDDCVDNHDMLVRIIVQLGHNMCKVIAEGVETEEQRLLVSEQ
CBEFGYLYSKPDLVDEEFNRDLQI"
complement(5356..5475)
/gene="VVA1131"
complement(5356..5475)
/gene="VVA1131"
/note="identified by Glimmer2"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC97157.1"
/db_xref="GI:37201335"
/translation="NSIPKPHNHSKGNHAKESWCKMNTQGRKFAVYTG"
5718..6596
/gene="VVA1132"
5718..6596
/gene="VVA1132"
codon_start=1
/transl_table=11
/product="putative transporter, DME family"
/protein_id="BAC97158.1"
/db_xref="GI:37201336"
/translation="MQSQHPRLIGASWMLTAGIAPAVINSLAQVASTIOFAPSTVAL
IQYGLIYILPYLKTIGIRQSLRTSHFPMHALRVFLSYIGQIMMALAVVPIVQ
IALMTSPFATIGSLILKEKVGAVRGATLFGFGAMIIEPMADDFHALTLFVG
AAPRWACVIMTKLSVNDPSTMVVYLLITTPNIMATIDPTWPSSTIYLLIFA
AGCMTLAQMALYKAYASDASFVQFPDAKIPINVLACMLVFGWVPPRMLWGAFTI
IASVARTHMETKTK"
6661..7266
/gene="VVA1133"
6661..7266
/gene="VVA1133"
codon_start=1
/transl_table=11
/product="glutaredoxin-related protein"
/protein_id="BAC97159.1"
/db_xref="GI:37201337"
/translation="MSEAIKITLYRWAGSMGPKVNIPOGECTLRDILTDPAEELN
GIPELEVDOMLSHMEPLKLGAMNAPILVVGKVSQCEALNRGLISVIREWTKR
DILKNIIVPGKATCPYCVAKOLDBAGIHGYHDVVKESALYKIPVKAIIIGKTI
PVTVPQIWLDPGATYIGCCDLTEYILKERGLDVIPNNVVEAN"
complement(7360..7494)
/gene="VVA1134"
complement(7360..7494)
/gene="VVA1134"
/note="identified by Glimmer2"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC97160.1"

| FEATURES | Source |
|------------|---|
| Db | 259437 ATGAAATTCATTATATACACGGTTTATATATGATGAGTTTAGTCATGCAACCTCTGACG 259439 |
| Qy | 61 CATGTCGTGCATAATTTGGGCTTTCGTACTCAACCACTTATGCTTAACAATCTGCTGATC 120 |
| Db | 259497 TTGGCGCTTAATAAAATTTGGGCTTTCATAACCAACCGCTCACTTACACACGCTCAATTT 259556 |
| Qy | 121 GATGATAGAGCCATTTTTCGCGCCCTTGACCGATGCTCATCTGACCTGCTTAATGCT 180 |
| Db | 259557 GATBAAABAAAAGTTTTCATTCATATGACCAAAAGCGCTCAGACGACACCGGAAATGTG 259616 |
| Qy | 181 TTAATCGCAACAAGTTTGGCGGAGTTGGTATCAAAAGTTATCTAAGATCGCGGACCG 240 |
| Db | 259617 TTGGTTGGCCATGTTTGGGAGGCTTGAATGATCAAAATATTTCTTGCCATCGTCGCT 259676 |
| Qy | 241 TCCTGTGAAAACCGCTTCCCATATGTCGTCGCAATCGGCTACCTTTGCAAGAGCTTCATT 300 |
| Db | 259677 ACCCTTGACAAATCTTCATATGTGTGGCCATATGGGTCAACGCTTAAGGCGCTTCCATC 259736 |
| Qy | 301 GTCATATAAAATTGACAAATTTAGTTTGGGATGAGCACTAGTAATTCAGCAAAATTTGG 360 |
| Db | 259737 GTATCAAAAATTCAGGAGGCTTGTATTTAGGTGCGATGCTTGGCAATTCGCCAAAACATGG 259796 |
| Qy | 361 TTAAABAAACACACACCAATCCGCTATCCACAAAATCAGGATGATTTGGCAGAACG 420 |
| Db | 259797 CTCATATGATGATGAGATGTTTGGATTTTCTCTCAAAAGCTTGGCAGATTTGCGAGC 259856 |
| Qy | 421 ATACCTTTAAGGCTGCGACAGCTTTTACTGCGCAGATCACTG---GATCCGATGTATCC 477 |
| Db | 259857 ATTCTATTTGGGTGCGAAGACCTTCTTACTGATGATGATACACACAAATCTCTGATGGACG 259916 |
| Qy | 478 GTCACAGTAGAAGAAACCAAAATATGCTGCGATGACAGATCATATGCGCA 526 |
| Db | 259917 GTCAAGGTGAGAGAGCAAAAGATATCAGGTATGACTGACATGATTTTGA 259965 |
| RESULT 3 | |
| AE004352/c | 10977 bp DNA linear BCT 10-OCT-2003 |
| LOCUS | Vibrio cholerae O1 biovar eltor str. N16961 chromosome II, section 9 of 93 of the complete chromosome. |
| DEFINITION | AE004352 AE003853 |
| ACCESSION | AE004352.1 GI:9657475 |
| VERSION | |
| KEYWORDS | |
| SOURCE | Vibrio cholerae O1 biovar eltor str. N16961 |
| ORGANISM | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio. |
| REFERENCE | 1. (bases 1 to 10977) |
| AUTHORS | Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Teitelin,H., Richardson,D., Ermolaeva,M.D., Vamathavan,J., Baas,S., Qin,H., Dragoi,I., Selters,P., McDonald,L., Ueberlack,T., Fleischman,R.D., Niernan,W.C. and White,O. |
| TITLE | DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae |
| JOURNAL | Nature 406 (6795), 477-483 (2000) |
| MEDLINE | 20406833 |
| PUBMED | 10952301 |
| REFERENCE | 2. (bases 1 to 10977) |
| AUTHORS | Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A., Gill,S.R., Nelson,K.E., Read,T.D., Teitelin,H., Richardson,D., Ermolaeva,M.D., Vamathavan,J., Baas,S., Qin,H., Dragoi,I., Selters,P., McDonald,L., Ueberlack,T., Fleischman,R.D., Niernan,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R., Mekalanos,D.J., Venter,J.C. and Fraser,C.M. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA |
| FEATURES | Location/Qualifiers |
| Source | 1..10977 |
| | /organism="Vibrio cholerae O1 biovar eltor str. N16961" |

gene /mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:243227"
/chromosome="II"
/note="biotype: E1 Tor"
complement (81..974)
/gene="VCA0090"
complement (81..974)
/gene="VCA0090"
complement (81..974)
/gene="VCA0090"
complement (81..974)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF96004.1"
/db_xref="gi:9657476"
/translation="MSLVNLLQLCLAAIWGGSFLFMRIAHSPGPAVLIEARVGPAL
LSLPIVAQLRRSLPIRQHPHPLITGLINAVPFLFAVALTLNVTLSILNSTA
INGAVIGFEGSTPLSRKAVAGLLIGSVGAVIVGMDVAIGHAALPMVCAALAA
YGLANNYQAQPOLAFENAGSMWACLWAPLPMFVLESTPSLEMGAVILLGVG
CGIAYVILFPLVYALIGAASITLSVFLPVGILMGVILDDPGLGNLPLFTLLVLA
TMLVYGFSTRNRRLSRQPPH"
1204..1500
/gene="VCA0091"
1204..1500
/gene="VCA0091"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF96005.1"
/db_xref="gi:9657477"
/translation="MARPKIPRIECPHSPACCFKNGVPIROLARVELAPDELALRL
VDQLGLOOQALQWQVSRQTLNLYKARPKVVDCLHQAALYIGALDNKSD"
complement (1671..2300)
/gene="VCA0092"
complement (1671..2300)
/gene="VCA0092"
complement (1671..2300)
/gene="VCA0092"
complement (1671..2300)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF96006.1"
/db_xref="gi:9657478"
/translation="MKIVILGLYNGSLWMPLAQRINKLQGYQEVISYNTLAIDKER
VQCIIDSLADRIYLVGSLIGLIMIGYLRSHSPSNVISHVALASPAGKSIYV
KIQGLGAGLNAHLVGLAHQDSMELPQRLGGLAGLRGPPILGSGMDGVYV
TYAEIQISGMDHLLHSHGTGLVYSKTAQIDVYFIRNQGFKKIPB"
2400..2573
/gene="VCA0093"
2400..2573
/gene="VCA0093"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF96007.1"
/db_xref="gi:9657479"
/translation="MEPCWMLRVNPIPTWFKYVNBDRCTLLPVPISRLAKGLEKVVH
DDCLFKSKSEODIM"
2801..3157
/gene="VCA0094"
2801..3157
/gene="VCA0094"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF96008.1"
/db_xref="gi:9657480"
/translation="MDVNIKSPRIPINDYALAQLERISTKRYTHILMHLDSSEAPF
GELQYISIAKMLSHRLKELIGVMVHQQPTLPVTVYIGLTSGRVYANGVLYW
IDHVENNERLEPLLEIN"

```

gene
3426. 4124
/gene="VCA0095"
CDS
3426. 4124
/gene="VCA0095"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA096009.1"
/db_xref="GI:9657461"
/transl_table="MYOMKLSKMLFSLVFLGILSLTYTSPPTPTNLPDRQEH
VAPKRPABIEIHSHPALVELESTENRVNVEQSEVQSEAPRVIEPQNLQVWLPI
NPTPELLADLKPKSOAHTDQALEFENIDQFVLMAGDSATLPKNGESIRVLEBACQ
OAGQIQTAGNFETQGTGTPVFTFGQSHMGFIQHPQSGIKIEGVGHQAWVYQVPS
HPEHDTOSHDBAH"
4167. 5390
/gene="VCA0096"
CDS
4167. 5390
/gene="VCA0096"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA096010.1"
/db_xref="GI:9657482"
/transl_table="MKKSLALAVSVGLCSPAIMATTEIDVLGYTPDTAGFRQEHV
AOMHVSANVANKVLKDSGLDIKVNLAATREVOYTOPGKKSQSEVLDAAPFNRIID
APADAVAYQGVAGDMVAIFRYLDVNNSPDYEROPNGSYISGLGAMVAASAMYPQ
NAKSVYSHSYINEGCAETFEIENHGNHGAHQAQVRELPNNHNGTVDVAGVIGIC
QRTITAYPHLFGVGSYKFSPPNOCSEAPCGKVDKANSVRAIGLIEPHIAQYVIGI
KPVVDNGFDTEPDNTNNVFTIKGPAIDPMKILTPIVSAQSSSTAVADIDTH
EYRGDISILFAPDGSYVWLKQANRYDRQSYNVQFTINDVDPASABGEWMLIODEP
GGLKLTINQFOITFP"
/notes="complement (5403. 6119)
/gene="VCA0097"
CDS
complement (5403. 6119)
/gene="VCA0097"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA096011.1"
/db_xref="GI:9657483"
/transl_table="MTVTDMTCRLAPKGIITLLVKSNNPNDCCGMAIPGQWFD
EULSAGQSPADKDPDSARRIQKQKHIPYFISDPVDANPKRDSGWSISISHYA
LHNSAVTQIEBAEMDKSRVNFALETILQGVAFVHQAQIQHAWKGRANAAVEYS
VVLFSLEFELVADIIIAAKFGVDVNRMTVARRLIDTGVISTKIAASSKGGKXP
ATVYRLASNEVYFOTCLAK"
5518. 7825
/gene="VCA0098"
CDS
5518. 7825
/gene="VCA0098"
/notes="similar to GB:U05568 SP:P18133 PID:147307 GB:U00096
PID:1651456; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="nicotinate phosphoribosyltransferase"
/protein_id="AA096012.1"
/db_xref="GI:9657484"
/transl_table="MNPRLFSPIHRIISLDDAYKINMQAIIHFPYDVSRYELIAR
SEEDASGLDAIROBIAGHGLTFSDDADHYLTQAPHLKATFLQSLRYFHVPOEY
ENGIYQGGKQOOLRISIRGSRDITLYETLWVAIVSEVRSQRMALVADPLKYLKT
KIDQCKAEIERGINNFSITENGTRRSPSOVQVRLACIKQEIPOWLTGTSNHFAR
EDPKIKTIGTAHEMWHQALINERDSQVALERPLTAPDGLAIAPDTDTLIDAFIN
DNRRLANAYDGVDRDSCGPPKGMKMTAHYQGLIDITPTLFTISDGLDPOALITC
BYPAGKVKISFGITPLINDLANWRKAKGVEYRPUSITVKLAECCGRVAAKISQPER
AMCEPDIFLANIKRRFNIELDVDAIIQELRHQKRSRPHYISA"
7856. 8896
/gene="VCA0099"
CDS
7856. 8896
/gene="VCA0099"
/notes="similar to GB:U00096 PID:1742678 PID:1742686
PID:1742694 PID:1742718; identified by sequence

```

```

similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, Gfc/Idh/MocA family"
/protein_id="AA096013.1"
/db_xref="GI:9657485"
/transl_table="MNPRLKVGIVIGYSAKTFHLPIDTLESLTSLAISSQOQV
OEDMPOIAYPDSACRLITSDYDVLITAPNNVHPIALALAEHGVIVKRPYTOI
EQRTELELRQOGLISVFNHNRMDGDLVTKALIEQGLGDKVRFESHPDRFVY
RQWRREQAGGGLIFDLAPHLDDQALVFGIPQSLADCKMRPDAITLDFDLOLY
YPOHVVRLHANTLSPSPNRYOVLGSLGVXKGLDPOBRLAKAGRPHPWQSQMP
EQQGILYHAGNENVTIEIGYGHYFTQVEAIEINRNASNPVSMSALOSIOLIELALE
SSAKGORLAVTL"
/notes="complement (8933. 9304)
/gene="VCA0100"
CDS
complement (8933. 9304)
/gene="VCA0100"
/notes="similar to GB:7226016; identified by sequence

Query Match 34.5%; Score 191.4; DB 1; Length 10977;
Best Local Similarity 60.5%; Pred. No. 2,6e-50;
Matches 315; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

1 ATGCAGATTATCTTGTTCATGCACTCTATATGCAATGCTGGTAAATGATCCGCTTACT 60
2300 ATGAAAGATTGATTAATACGACAGGCTTGATATGCAATGCTTGGTAAATGCAACCTTCGCG 2241
61 CATGCTGATTAATGCTGGTAAATGCTGATCAACCAATTAGCTACAACTGCTGCTATC 120
2240 CAACGTCATTAATACCTGATATACAGACGAAAGTGTACAGCTACAACTTATGCAATC 2181
121 GATGATGAGCCATTTTTCGCGCCCTTGACCAATGCTCACTGATGCTGCTGCTAAATGCT 180
2180 GATGATGAAAGATTTTTCAGACATTAAGACAGGCACTAGACAAATGCTGATCAACGTTG 2121
181 TTATGCGACACAGTTTGGCGGATGCTGATGATGATGATGATGATGATGATGATGATGATG 240
2120 TTGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2061
241 TCCGTGAAACCCCTCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
2060 TCACCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2001
301 GTCATTAATTAATGCAATTAATGCTTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
2000 GTCCTTAATTAATGCAATTAATGCTTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1941
361 TTAAGAAACACGACGACGATCCCGCTATCCCAAAAATCAGGACGATATTGACAGGACG 420
1940 TTGCAATGATCAAGATGATGCTGGAAATTGCCAACAAGGCTGGTTGATTCAGAGAAC 1881
421 ATACCTTTAGGCTGGCGACCTTTTACTGCGGATCCACTGATCCGATGCTACCGTTC 480
1880 TTGGTTTCGATTTTGGCCCAATCTTACCTCGGCTCAGGAATGTGTATGACAGCGTC 1821
481 ACAGTAGAAGAAACAAATATGCTGCGATGACGATCATAT 521
1820 ACGGTTGCCAGACTCAATCTCAGGAATGACGATCATCT 1780

RESULT 4
AP005088 339350 DNA linear BCT 05-MAR-2003
LOCUS Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 5/6.
DEFINITION AP005088 BAO00032
ACCESSION AP005088.1 GI:28809504
VERSION
KEYWORDS
SOURCE
ORGANISM
Vibrio parahaemolyticus
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
1 Nesu,H., Iida,T., Sugahara,T., Yamachi,Y., Park,K.S., Yokoyama,K.,

```

TITLE Makino, K., Shingawa, H. and Honda, T.
A filamentous phage associated with recent pandemic Vibrio
parahaemolyticus O3:K6 strains
JOURNAL J. Clin. Microbiol. 38 (6), 2156-2161 (2000)
MEDLINE 20295086
PUBMED 10834969
REFERENCE 2
AUTHORS Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,
Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A.,
Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shingawa, H.,
Hattori, M. and Iida, T.
TITLE Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae
JOURNAL Lancet 361 (9359), 743-749 (2003)
MEDLINE 22508454
PUBMED 12620739
REFERENCE 3 (bases 1 to 339350)
AUTHORS Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T.,
Honda, T., Shingawa, H., Hattori, M. and Iida, T.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken-gen-info.osaka-u.ac.jp/
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
COMMENT This clone was isolated from a patient presenting with acute
gastroenteritis.
FEATURES
SOURCE Location/Qualifiers
1. .339350
/organism="Vibrio parahaemolyticus"
/mol_type="genomic DNA"
/strain="O3:K6"
/sub_strain="RIMD 2210633"
/db_xref="taxon:670"
/chromosome="2"
64. .1428
/gene="VP1137"
64. .1428
/gene="VP1137"
/note="similar to GB:AA119314.1 (AE008712) percent
identity 76 in 437 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="cytochrome BD2, subunit I"
/protein_id="BAC62480.1"
/db_xref="GI:28809505"
/translation="MFGDAPMLARIQPAFTVSHITPAITTGATYLAVALBGLMK
TRNDYQLYHFMKIFAVNMGVAVGLVMAVQFTNWSGSDPASITGPIITVEV
LTAFFLEAGPLGVLFGMRVGEKLFHFAATSMVALGTIISTFWILASNMWQPOGE
IVDRVVTDFWPAIVNPSFVRLAHSAVAFASSALFVGAAMHLIRNOSTAVYT
MFMSLIGLIVLPAIVAGIDVGHNLTEHOPALIEGHMNSDCKPRLIFGAP
NMBOERDVALEIEVLGSLIRHSITRIPALDKPKEERNSPIVMSQPRIVAGL
LMIPQSYSMYLRKNTLYTSRMTFKSLSPKBSGLPAIILAGPTTGVGQPPVYVGV
QKTRDAVASGDLQMSISLDFLVVSLVFGFYTHIQIKGPDALIEHDEHDMITV
SGRI"
1457. .2464
/gene="VP1138"
1457. .2464
/gene="VP1138"
/note="similar to GB:AA119315.1 (AE008712) percent
identity 67 in 336 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="cytochrome BD2, subunit II"
/protein_id="BAC62481.1"
/db_xref="GI:28809506"
/translation="MHPDLSYTPAIIIVPATIMTINDGPDIGITIMPPIKQKQD
VMVNSVAVWDGNTWIVLGSALFSGAPPAVAVITATITPILMLAIFGVAFS
FRFQLEHLKFMRSPIVLSILTFPGGVAGVAVIQGFTVENRVPVGSQDLWLVAFS
IFQIGLVATVALGSTMILKTEGALONTMYRPTNTLLAMISALIVSAMPIAYP

gene 2469. .2627
/gene="VP1139"
2469. .2627
/gene="VP1139"
/note="unknown"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC62482.1"
/db_xref="GI:28809507"
/translation="MIRSEYENQKLVPPVGLDWDVLGGQCASSGVFINAVSHDRCR
VILTRHP"
complement (2684. .3439)
/gene="VP1140"
complement (2684. .3439)
/gene="VP1140"
/note="similar to GB:AA96301.1 (AE004375) percent
identity 62 in 250 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC62483.1"
/db_xref="GI:28809508"
/translation="MKAQIYDVFCDELAQGNPCGVVLEINWLSSELIQTTHQAQ
VTSFVKSQSDHIRMFTLAGINLCGSGLDAGALISKHONVILQSSGYITVS
ELIDGMVRIYMPGMAKPAHPLGISGFENRPIQDAFTTDLVNVSEDDVVCYCNPD
ELIKIDFPAIVTAQSSDSGVLYRFAPKIGISDLATGSAQCSLAWYWEKLGKKEK
LTAQLSRSGYEVAVRGSRTIVLSAQVSRV"
complement (3534. .3911)
/gene="VP1141"
complement (3534. .3911)
/gene="VP1141"
/note="similar to GB:AA884740.1 (AE000810) percent
identity 47 in 116 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="gamma-carboxymuconolactone decarboxylase"
/protein_id="BAC62484.1"
/db_xref="GI:28809509"
/translation="MENTRFQGTGEQLSKIDGAGQVIESINDICPDIAKFTIYRP
GDIYVRGDLKSRRIATVSALTAMGNCPTPOLKVLHAAALNVGTEBEIKETIIGMSV
YAGPPAALNGMAFKEVLSEROV"
complement (3970. .4326)
/gene="VP1142"
complement (3970. .4326)
/gene="VP1142"
/note="similar to GB:AAK76923.1 (AE001438) percent
identity 48 in 109 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative transcriptional regulator, MerR family"
/protein_id="BAC62485.1"
/db_xref="GI:28809510"
/translation="NMKKESSAVGLSYYTLRYEKIGLLKVRHNSGHRVYTRDI
DMVNFILKXKGMLEIEIOYASLRSGSTTDROKLLAEHRNLLIEHRIQGNELH
KLEERKINDLYSGKRV"
4561. .5874
/gene="VP1143"
4561. .5874
/gene="VP1143"
/note="similar to EMB:CB72498.1 (AL139074) percent
identity 45 in 406 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative molybdenum containing oxidoreductase"

* 20914 22022: contig of 1109 bp in length
* 22023 22122: gap of 100 bp
* 22123 22324: contig of 1102 bp in length
* 22324 22324: gap of 100 bp
* 22325 22496: contig of 1172 bp in length
* 22497 22496: gap of 100 bp
* 22565 22564: contig of 668 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 1384 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 4950 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 5840 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 14931 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 12442 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 8682 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 20389 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 39537 bp in length
* 22565 22564: gap of 100 bp
* 22565 22564: contig of 44560 bp in length.

FEATURES
Location/Qualifiers
1..178879
source

misc_feature
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_id="RP24-440120"
/clone_id="RP24-440120"
1..416
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
517..1177
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
1278..1966
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
2067..2713
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
2814..3516
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
3617..4331
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
4432..5089
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
5190..5804
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
5905..6568
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
6669..7349
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
7450..7732
/note="assembly_fragment"

Query Match
Best Local Similarity 5.7%; Score 37.4; DB 2; Length 178879;
Matches 71; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 409 ATTGAGAGGATACCTTTAGGCTGCGAGCCTTTTACTGCGGATCACTGAGCTCC 468
DB 17737 ATGCGCGGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 17678
QY 469 GATGATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 528
DB 17677 GATGATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 17618
QY 529 TCCACCA 535
DB 17617 CCGGCA 17611

RESULT 7

AX655393
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Location/Qualifiers
1..2000
source

ORIGIN

Query Match
Best Local Similarity 7.7%; Score 37.2; DB 6; Length 2000;
Matches 39; Conservative 231; Mismatches 234; Indels 0; Gaps 0;

QY 30 TATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 89
DB 523 TATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
QY 90 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 149
DB 583 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 642
QY 150 CCGATGCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 209
DB 643 SRSRWSKSTYCYRKRKSKSTYCYRKRKSKSTYCYRKRKSKSTYCYRKRKSKSTYCY 702
QY 210 GATCAACATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 269
DB 703 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 762
QY 270 CCGATGCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 329
DB 763 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 822
QY 330 GATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 389
DB 823 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 882
QY 390 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 449
DB 883 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 942
QY 450 GCGGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 509
DB 943 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 1002
QY 510 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 533
DB 1003 TCAACCATTTAGTCACTCACTGCTGATGATGATGATGATGATGATGATGATGATGAT 1026

RESULT 8
AX251505/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
AX251505
Sequence 473 from Patent WO0168912.
AX251505.1 GI:15984928
synthetic construct

ORGANISM synthetic construct
artificial sequences.

REFERENCE
AUTHORS
TITLE
JOURNAL
Epidemiology AG (DE)
Patent: WO 0168912-A 473 20-SEP-2001,
Oncogenes
Location/Qualifiers
1..26997
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 6.6%; Score 36.8; DB 6; Length 26997;
Best Local Similarity 47.1%; Pred. No. 5.6;
Matches 113; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 277 TCACCTTGCAGAGGCTTCATTTGCAATAAATTGAGCAATTAGGTTAGGGGCGCA 336
Db 13416 TCTCCATGGAACACTACCACTATTAACAACTTAACCTTAATACCTTAATTC 13357
QY 337 CTAGGTAATTCAGCAAGATTGGGTTAAAGACGACGAGCAATCCCGTATCCACA 396
Db 13366 CTACACAACTTATCACTTCAGCACTAATACCACTCGAACAATCACTTTAAACAA 13297
QY 397 AAATAGGACGATTTGCGAGCAACGATTAAGGCGTGGCGACCTTTACTGGCGAT 456
Db 13296 AAAAAAATTAACCTACCAAAACGCTACCTATTAATCAAAATATCCCGACCTTAC 13237
QY 457 CCAGTGCAGCTCCGATGCTACCGTCAAGTAGAAGAAACCAATATGCGATGACAGAT 516
Db 13236 CCACCAACACCAATTAACCTCCCACTATTAACCAAAACGACAACTTAACAAAT 13177

RESULT 9
AL592225 183397 bp DNA linear ROD 10-JUN-2002
LOCUS Mouse DNA sequence from clone RP23-278M14 on chromosome 1, complete
DEFINITION
ACCESSION AL592225
VERSION AL592225.17 GI:21711841
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 183397)
Blakey, S.
Direct Submission
Submitted (10-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail: enquires@sanger.ac.uk
humquerry@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Jul 8, 2002 this sequence version replaced gi:21615572.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-278M14 is from the RPCT-23 Mouse PAC library
constructed by the group of Peter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1..183397
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP23-278M14"
/clone_1lb="RPCT-23"

ORIGIN

Query Match 6.6%; Score 36.8; DB 10; Length 183397;
Best Local Similarity 56.7%; Pred. No. 7.6;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 288 AGGAGCTTCATTTGCAATAAATTGAGCAATTAGGTTAGGGGCGCACTAGTAATTC 347
Db 58507 AGCAGCTTAATGAGCATTAATTAACAGACGAGTGGTCTTCTTGATGACAT 58566
QY 348 AGCAAAATTTGGTTAAAGAACGACGACGACGATCCGCTATCCACAAATTCGCGC 407
Db 58567 GGCTTAATCAGCTTAAGAACACACAGAGGAGAGTCTCTGTAGGCAAACTCAGGTG 58626

RESULT 10
AC130676/c 209061 bp DNA linear HTG 10-MAR-2003
LOCUS Mus musculus clone RP23-167L15, WORKING DRAFT SEQUENCE, 13
DEFINITION
ACCESSION AC130676
VERSION AC130676.3 GI:28894622
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209061)
Birren, B., Nussbaum, C. and Lander, E.
Mus musculus, clone RP23-167L15
Unpublished
2 (bases 1 to 209061)
Birren, B., Nussbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,
Barra, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, V., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galasani, J.,
Gargana, S., Gird, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhthang, P., Pierre, N., Raymond, C., Retta, R., Rice, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuppback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zemlek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE

Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 209061)

AUTHORS
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Bana, N., Bastien, V., Bloom, T., Boggislaw, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamei, A., Karas, A., Kelis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melgrim, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhag, P., Pierre, N., Rachupka, A., Ramsamy, U., Raymond, C., Rettig, R., Rise, C., Rogov, P., Roman, V., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
 Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 10, 2003 this sequence version replaced g1:28269519.
 All repeats were identified using RepeatMasker:
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

JOURNAL
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

COMMENT
 Project Information
 Center project name: 167.L.15
 Center clone name: 167.L.15
 Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 205696 bases at least Q40
 Consensus quality: 205796 bases at least Q30
 Consensus quality: 207355 bases at least Q20
 Insert size: 21000; agarose-ftp
 Insert size: 207861; sum-of-contents
 Quality coverage: 8.4 in Q20 bases; sum-of-contents
 Quality coverage: 8.5 in Q20 bases; sum-of-contents

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 20644: contig of 20644 bp in length
 20645 20744: gap of 100 bp
 20745 21744: contig of 1000 bp in length
 21745 21844: gap of 100 bp
 21845 23426: contig of 1582 bp in length
 23427 23526: gap of 100 bp
 23527 25861: contig of 2235 bp in length
 25862 25863: gap of 100 bp
 25864 27654: contig of 1793 bp in length
 27655 27754: gap of 100 bp
 27755 33828: contig of 6074 bp in length
 33829 33928: gap of 100 bp
 33929 40078: contig of 6150 bp in length
 40079 40178: gap of 100 bp
 40179 89252: contig of 49074 bp in length
 89253 89353: gap of 100 bp
 89353 110083: contig of 20731 bp in length
 110084 110183: gap of 100 bp

FEATURES
 source
 * 110184 133812: contig of 23629 bp in length
 * 133813 133912: gap of 100 bp
 * 133813 159643: contig of 25731 bp in length
 * 159644 159743: gap of 100 bp
 * 159744 205188: contig of 45445 bp in length
 * 205189 205288: gap of 100 bp
 * 205289 209061: contig of 3773 bp in length.
 Location/Qualifiers
 1..209061
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-167L15"
 /clone_1lb="RP23-167L15"
 1..20644
 /note="assembly_fragment"
 clone_end:826
 vector_side:left
 20745..21744
 /note="assembly_fragment"
 21845..23426
 /note="assembly_fragment"
 23527..25761
 /note="assembly_fragment"
 25862..27654
 /note="assembly_fragment"
 27755..33828
 /note="assembly_fragment"
 33829..40078
 /note="assembly_fragment"
 40179..89252
 /note="assembly_fragment"
 89353..110083
 /note="assembly_fragment"
 110184..133812
 /note="assembly_fragment"
 133913..159643
 /note="assembly_fragment"
 159744..205188
 /note="assembly_fragment"
 205289..209061
 /note="assembly_fragment"
 clone_end:17
 vector_side:right

ORIGIN
 Query Match 6.6%; Score 36.4; DB 2; Length 209061;
 Best Local Similarity 50.6%; Pred. No. 10;
 Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Gy 193 AGTTGGGGGATGGATGATCAAGCTTATCTGAATCGGCGGACCGTCTGTGAACC 252
 Db 137780 ATTATAGCGGAAGATCAATATGTTAGCTTTAGCAAAAGCTATGCTGTGAGA 137721
 253 CTCCTCCATGCTGTCGCATCGGCTCACTTTCAGAGAGCTTCAATGTCAATAAAT 312
 Db 137720 CAGTTCACCTCTTGTCTTGATCTTAACATTTCCAGAGAACCTTCATATCAAAAT 137661
 Gy 313 GAGCAATAGCTTTAGGGGCGGACATAGTAATTCAGCAAGAAATTTGGGTTAAA 366
 Db 137660 CAGTCATCTGACCTAGACTGACCTGAGTAGTCTTCATGCAAGAGATAGAA 137607

RESULT 11
 AC130294 54762 bp DNA linear HTG 19-SEP-2002
 AC130294
 DEFINITION Homo sapiens chromosome 11 clone RP13-57287 map 11, LOW-PASS
 SEQUENCE SAMPLING.
 AC130294
 AC130294.3 GI:23196586
 VERSION
 KEYWORDS HTG; HTGS PHASEO.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 54762)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP13-572P7
Unpublished
2 (bases 1 to 54762)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukigalter,B.,
Camara,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatza,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meltrin,J., Meneus,L., Mihova,T., Menga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 54762)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukigalter,B.,
Camara,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatza,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meltrin,J., Meneus,L., Mihova,T., Menga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (19-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 19, 2002 this sequence version replaced gi:22532570.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L27832
Center clone name: 572_F_7

* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

* be preserved.
1
706 705: contig of 705 bp in length
806 805: gap of 100 bp
1513 1512: contig of 707 bp in length
1613 1612: gap of 100 bp
2330 2329: contig of 717 bp in length
2430 2429: gap of 100 bp
3149 3148: contig of 719 bp in length
3249 3248: gap of 100 bp
3955 3954: contig of 706 bp in length
4055 4054: gap of 100 bp
4754 4753: contig of 699 bp in length
4854 4853: gap of 100 bp
5564 5563: contig of 710 bp in length
5664 5663: gap of 100 bp
6373 6372: contig of 709 bp in length
6473 6472: gap of 100 bp
7182 7181: contig of 709 bp in length
7282 7281: gap of 100 bp
7974 7973: contig of 693 bp in length
8074 8073: gap of 100 bp
8795 8794: contig of 719 bp in length
8794 8793: gap of 100 bp
8894 8893: gap of 100 bp
9602 9601: contig of 708 bp in length
9702 9701: gap of 100 bp
10388 10387: contig of 686 bp in length
10488 10487: gap of 100 bp
11192 11191: contig of 704 bp in length
11292 11291: contig of 716 bp in length
12007 12006: contig of 716 bp in length
12107 12106: gap of 100 bp
12799 12798: contig of 692 bp in length
12800 12799: gap of 100 bp
12800 12799: gap of 100 bp
13594 13593: contig of 695 bp in length
13694 13693: gap of 100 bp
14518 14517: contig of 724 bp in length
14518 14517: gap of 100 bp
15237 15236: contig of 719 bp in length
15337 15336: gap of 100 bp
15338 15337: gap of 100 bp
16048 16047: contig of 711 bp in length
16049 16048: gap of 100 bp
16149 16148: gap of 100 bp
16853 16852: contig of 705 bp in length
16854 16853: gap of 100 bp
16954 16953: contig of 680 bp in length
17634 17633: gap of 100 bp
17734 17733: gap of 100 bp
18432 18431: contig of 699 bp in length
18532 18531: gap of 100 bp
19244 19243: contig of 712 bp in length
19344 19343: gap of 100 bp
20060 20059: contig of 716 bp in length
20161 20160: gap of 100 bp
20867 20866: gap of 100 bp
20968 20967: contig of 707 bp in length
21672 21671: gap of 100 bp
21772 21771: contig of 705 bp in length
21773 21772: gap of 100 bp
22489 22488: contig of 717 bp in length
22589 22588: gap of 100 bp
23301 23300: contig of 712 bp in length
23401 23400: gap of 100 bp
24115 24114: contig of 714 bp in length
24215 24214: gap of 100 bp
24932 24931: contig of 707 bp in length
24933 24932: gap of 100 bp
25022 25021: gap of 100 bp
25724 25723: contig of 702 bp in length
25824 25823: gap of 100 bp
26319 26318: contig of 695 bp in length
26520 26519: gap of 100 bp
27326 27325: contig of 707 bp in length
27427 27426: gap of 100 bp
28149 28148: contig of 723 bp in length
28249 28248: gap of 100 bp
28957 28956: contig of 708 bp in length
29057 29056: gap of 100 bp

```

* 25058 29771: contig of 714 bp in length
* 29772 29871: gap of 100 bp
* 29872 30595: contig of 724 bp in length
* 30596 30695: gap of 100 bp
* 30696 31407: contig of 712 bp in length
* 31408 31507: gap of 100 bp
* 31508 32219: contig of 712 bp in length
* 32220 32319: gap of 100 bp
* 32320 33049: contig of 730 bp in length
* 33050 33149: gap of 100 bp
* 33150 33858: contig of 709 bp in length
* 33859 33958: gap of 100 bp
* 33959 34654: contig of 696 bp in length
* 34655 34754: gap of 100 bp
* 34755 35476: contig of 722 bp in length
* 35477 35576: gap of 100 bp
* 35577 36298: contig of 722 bp in length
* 36299 37109: contig of 711 bp in length
* 37110 37209: gap of 100 bp
* 37210 37900: contig of 691 bp in length
* 37901 38001: gap of 100 bp
* 38001 38695: contig of 695 bp in length
* 38696 38795: gap of 100 bp
* 38796 39497: contig of 702 bp in length
* 39498 39597: gap of 100 bp
* 39598 40284: contig of 687 bp in length
* 40285 40384: gap of 100 bp
* 40385 41092: contig of 708 bp in length
* 41093 41192: gap of 100 bp
* 41193 41894: contig of 702 bp in length
* 41895 41994: gap of 100 bp
* 41995 42682: contig of 688 bp in length
* 42683 42782: gap of 100 bp
* 42783 43498: contig of 716 bp in length
* 43499 43598: gap of 100 bp
* 43599 44317: contig of 719 bp in length
* 44318 44417: gap of 100 bp
* 44418 45139: contig of 722 bp in length
* 45140 45239: gap of 100 bp
* 45240 45946: contig of 707 bp in length
* 45947 46046: gap of 100 bp
* 46047 46753: contig of 707 bp in length
* 46754 46853: gap of 100 bp
* 46854 47560: contig of 707 bp in length
* 47561 47660: gap of 100 bp

Query Match      6.5%; Score 36.2; DB 2; Length 54762;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 305 ATAAATTGACATTAGTTAGGGTGGCACTAGTAATTCAGCAGAAATTGGGTTAA 364
    |||
Db 26121 ATGAGAGAGATGATGATATAGACATGGAAGAAAGAAAGAAAGTTGGTTAA 26180

QY 365 AAGAAACACGACGACGAATCCCGCTATCCACAAAATCAGCGCATATGAGAGAGATAC 424
    |||
Db 26181 AAAAGGAGGAGGAGGAGGAGAAACCCATCTCTGACAGCCCTGAGTCCCTGAACTAGCT 26240

QY 425 CTTTAAAGGCTGCGACGCTTTTACTGCGGATCCACTGATCCGATGTGACCGTCA 482
    |||
Db 26241 CTGTAGCCGAGGCTGCGCTTGAATCTCAGAGATCCATCGCTCCGCTCGCTCGC 26298

RESULT 12
AP003850 101342 bp DNA linear PLN 23-JUL-2003
LOCUS AP003850
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
ACCESSION AP003850
VERSION AP003850.2 GI:22415917
KEYWORDS Oryza sativa (japonica cultivar-group)

```

```

ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GAS) genomic DNA, chromosome 7, BAC
clone: OJ1793_E11
Published only in Database (2001)
2
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468
On Aug 21, 2002 this sequence version replaced gi:14595195.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), GeneMark.hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerX
(http://www.tigr.org/edb/glimmerx/glmr_form.html), RiceHM
(http://rgp.dna.affrc.go.jp/RiceHM/), SplicePredictor
(http://biinformatics.iastate.edu/cgi-bin/sp.cgi), slm4
(http://globin.cse.psu.edu/html/docs/slm4.html), gap2
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI Nonredundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequence database at RGP or DBP. Protein homologues of the coding
regions were searched against NCBI Nonredundant Protein database
with BLASTP. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DBP accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DBP accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRGSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from -21M3 to M13rev of the BAC
clone. This sequence of OJ1793_E11 clone has an overlap with
OSUNBA0036M16 (DBP: AP005103) clone at 5' end and with OJ1657_A07
(DBP: AP003910) at 3' end. The sequence was generated by combining
Monanto and RGP-Japan sequencing data. Detailed information on
overlapping and assembly quality together with annotation of this entry
is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.
Location/Qualifiers
1..101342
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="OJ1793_E11"
/complement(1..1396)
/gene="OJ1793_E11.101"
/complement(1..1396)
/gene="OJ1793_E11.101"
/note="3' LTR"
/complement(3384..8953)
/gene="OJ1793_E11.101"
/complement(3384..8953)
/note="probably inactive due to including stop codon(s) in
CDS
misc_feature
gene
LTR
gene
source
FEATURES

```


REFERENCE
AUTHORS

1 (bases 1 to 271370)

Ratus.

Muzny, D., Maize, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bismail, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Garcia, R., Garcia, A., Garner, I., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kiyat, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenz, B., Lounsbury, L., Loulgesed, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, L., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, N., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarnpungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polinder, A., Popovic, D., Prims, E., Pu, L., R., Rivas, M., Quirz, J., Rachlin, E., Reeves, K., Register, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Soza, J., Steele, M., Strong, R., Sutton, A., Syarik, A., Taylor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umant, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

Weinstock, G., and Gibbs, R. A.

COMMENT

REFERENCE
AUTHORS
JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

On Sep 23, 2002 this sequence version replaced gi:21737475. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both ends and sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GPRD
Center clone name: CH230-71F9

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 186807 bases at least Q40
Consensus quality: 192016 bases at least Q30
Consensus quality: 196259 bases at least Q20
Estimated insert size: 199140; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

----- NOTES: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 48361: contig of 48361 bp in length
* 48362 48461: gap of unknown length
* 48462 68179: contig of 19718 bp in length
* 68180 68279: gap of unknown length
* 68280 271370: contig of 203091 bp in length.

FEATURES

source

1..271370
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-71F9"
18147..19710
/note="wgs contig"
misc_feature
48462..50098
/note="wgs contig"
misc_feature
58905..60802
/note="wgs contig"
misc_feature
68280..70194
/note="wgs contig"

ORIGIN

Query Match 6.5%; Score 36; DB 2; Length 271370;
Best Local Similarity 58.3%; Pred. No. 15;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

RESULT 14

CEW0647 27243 bp DNA linear INV 10-DEC-2003
LOCUS CEW0647
DEFINITION Caenorhabditis elegans cosmid W0647, complete sequence.
ACCESSION 278066
VERSION 278066.1 GI:1487930
KEYWORDS HTG; Histone H4 like.
SOURCE Caenorhabditis elegans

ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.

REFERENCE 1
AUTHORS none.
TITLE Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The *C. elegans* Sequencing Consortium.
AUTHORS 2 (bases 1 to 27243)
JOURNAL Ainscough, R.
TITLE Direct Submission
REMARK Submitted (09-AUG-1996) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
j@sanger.ac.uk or r.w.nematode@wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
COMMENT Current sequence finishing criteria for the *C. elegans* genome
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=W06A7)
name=W06A7
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone W06A7.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone W06A7 is at 1 in this sequence. The true
right end of clone W06A7 is at 9975 in
Sequence Z78067.
The true left end of clone ZC412 is at 2738 in this sequence. The
true right end of clone C48G7 is at 10917 in this sequence. The
start of this sequence (1..103) overlaps with the end of sequence
Z78061.
The end of this sequence (2738..27243) overlaps with the start of
Sequence Z78067.
FEATURES
SOURCE Location/Qualifiers
1..27243
/organism="Caenorhabditis elegans"
/mol_type="Genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="W06A7"
join(4610..4719,4772..4870,5394..5604,5655..5725,
6067..6106)
/gene="W06A7.2"
join(4610..4719,4772..4870,5394..5604,5655..5725,
6067..6106)
/gene="W06A7.2"
/standard_name="W06A7.2"
/codon_start=1
/product="Hypothetical protein W06A7.2"
/protein_id="CA801521.1"
/db_xref="GI:3880493"
/translation="MKTKIVDGVVPRVSRAPVPCFCLPAYMGVWPAKVVLLIG
ALLFFPVITILLMLCTIAIECGVASALIPALITITIGILLFRCGAAHLLDHG
EIPFKITVITITVMDPESHEDEIRLFTVHRBQVFNWQGFQGPENLSWQASNP
YPIQHTKNCLEEQPY"
complement(9699..9803)
/gene="W06A7.5"
complement(9699..9803)
/gene="W06A7.5"
/standard_name="W06A7.5"

gene /note="cDNA EST AM057186 comes from this gene"
/codon_start=1
/product="Hypothetical protein W06A7.5"
/protein_id="CA801519.1"
/db_xref="GI:3880491"
/translation="MNAIYAVLVASTLAYTAMAWIGSIEANEDLI"
join(10078..10235,10496..10565,10609..10746,11390..11511,
11614..11806,11854..11946)
/gene="W06A7.4"
join(10078..10235,10496..10565,10609..10746,11390..11511,
11614..11806,11854..11946)
/gene="W06A7.4"
/standard_name="W06A7.4"
/note="cDNA EST CE8942F comes from this gene
cDNA EST CE8942F comes from this gene
cDNA EST yk1089h08.3 comes from this gene"
/codon_start=1
/product="Hypothetical protein W06A7.4"
/protein_id="CA801524.2"
/db_xref="GI:15718229"
/translation="MSHVGMAMFMLPSSISGLTGMATISRFGLPGVPAVPSILLNY
NGIILLVGSSTIPEIKVNVGASDVTYLPKMKLKHOTNTEDVTSYMKVIAMI
GVIVGLIGWTTPTVITLMTRHDVPSKSTSPSHSPFTGVMVSVIIFYAC
PRKRPSTPRIVIPSMSCVFESGMAACFFIANEQSPTISYICMARGMTISAW
SVYFRHSRGNLLDGLRATGFLVGVLTASRVOL"
complement(join(12347..12413,12490..12861,12918..13068,
13512..13590,14170..14538,16386..16466,16510..17323,
19088..19168,20833..21123,21182..21300,21521..26999))
/gene="ret-1"
complement(join(12347..12413,12490..12861,12918..13068,
16386..16466,16510..17323,19088..19168,20833..21123,
21182..21300,21521..26999))
/gene="ret-1"
/standard_name="W06A7.3c"
/note="cDNA EST yk7f10.3 comes from this gene
cDNA EST yk7f10.5 comes from this gene
cDNA EST yk285f11.5 comes from this gene
cDNA EST yk608b12.5 comes from this gene
cDNA EST yk634c7.5 comes from this gene"
/codon_start=1
/product="C. elegans RET-1 protein (corresponding sequence
W06A7.3c)"
/protein_id="CA851467.1"
/db_xref="GI:5596654"
/translation="MEAPAVVIGLSDGNCNLRPALETSDQSSSENKPKDASNEVET
EPVSAKYALSSDGEENIKISANSVSPDADPRTGDEPTITILPSETTVT
QLEDLSTIEKQNEVESIKSEKRVGLIEBSGDTFRVTSGLAPDEDEARITTD
VPLSAGDIPSENEIKVSAADVVGLIEYVIGNPNAVNDIIPNFTPEVANDT
VTFSTVTAESAIPVVELEPIGDEYEFQEVENFSEPSDINIEESGAEVLENNM
FTPLDELPOOKILNEKAHEIIEASGDEPKDKHFPENONVWESRVEYVFI
GLESGLVLTGAVSDSVANNVKNESPDITSLSGDELKLVARELITTESKQAS
TDVPSKRTVSDVIGLEAGDKIVSNVSNMCKPDSGLLEQADVPELPINSE
ETVAVKATESCDHVDSQTLERASLSLEDVMSFEVLTISQTLSDVLPATSEBDS
IPVATVEETSEKVLSEEDVSVLEEDVLRVONKPSBETVAVDASKBGS
SDSPSRATETFMKLVITENTLIPAGDKSEBERIKDRENTISQPKEDDLENN
DPDDTIVIEKYSMAESLPIEAVISTDGGTSDQPAONAIIPDESETVDSQTEIF
TDDVYKSKENTPKAENDEITINYVJGEEGEDNAEKREAVSPDETSEIKODLEN
ENGASGPNNVQVPEAAQEDPDPEVETVETSENNPKAPTEDDNATEITIGLE
TTBAFGDAEHSYLDANIEKIVANADEPLPDELVSTIEERPEEVAPESTGEDNIFR
RDRRTVSLTGTDDNAPLOVIFVGGDGNIPVANNQERTSHNELTESDKSEBRLTK
NEEDVDQPIQSEBPLTQGEBSIGNKIVAVVGSVLGGAVTIPGLVASENEBAHA
DREVEETSDTRDREBETPVSKLSVMEVNLPSNDNIPKAVSENVLPVNTGLO
ESKENDPAPATAHSGKNRNDKTTRDRDEEITLKLVAENALPTGVTSQFTE
VSAPOAELETVVHAQNDTPSEVDEAPESAGGVTEKTSMTESILIPVAPQPE
NASHIIDETGASRIKDDONQPEEFSHGQKVSAPPOESAPEFLEAKDDQKE
TTENSEDAKRETVMEKYSVLENTIPYAVLPSTGTVPKAPKAPITIVQSSDEPDSI
TSQPELVPEISERVSDPEIFQRYKASSTIEPTQKTRPAPKAPITIVQSSDEPDSI
ANVIDELVHEDKRVPEVTANISVSASNTIDSTTAAVPRKTVESQLOVATYER
LESAPESASAIPVQPELKEVEVQPDLSQSPAPHKIIDAFNIPKQHEVGVNDYV
FGTESSESQKADQNOENQEDDVAAEINFPRIQWRBDEVISQSLKSLVAEVCIT
DVDADVNOEESTLKITLVPSGLPDEFTNDPPIIVLIPIMAPATVLEMY
EMILADAVESEMEVTESEISMAPVSESTQDIPPLADLKIPVEDDKETPEEP
VVPQGVQRITIPIVEQAPTIPQRPAPKPKSELPRVAPLDDSKSRVAFAPLNTLGR

Mon May 3 09:19:49 2004

us-10-603-260-4.rge

Page 19

Search completed: April 30, 2004, 15:29:14
Job time : 2249.64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 30, 2004, 06:19:01 ; Search time 227.167 Seconds
(without alignments)
10378.910 Million cell updates/sec

Title: US-10-603-260-4

Sequence score: 555
1 atgcagattatcttctgttca.....cttcatacagaaatgcgtt 555

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2003as:*
8: Geneseq2003cs:*
9: Geneseq2003ds:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 76 | 13.7 | 605 | 2 | AAT79331 |
| 2 | 37.2 | 6.7 | 2000 | 7 | ADA71938 |
| 3 | 36.8 | 6.6 | 26997 | 4 | AAG46747 |
| 4 | 35.6 | 6.4 | 115218 | 7 | ACA64845 |
| 5 | 34.4 | 6.2 | 12877 | 2 | AAT13834 |
| 6 | 34.4 | 6.2 | 2277 | 2 | AAV05370 |
| 7 | 33.2 | 6.0 | 2000 | 7 | ADA71938 |
| 8 | 32.8 | 5.9 | 948 | 7 | ACA27498 |
| 9 | 32.4 | 5.8 | 91080 | 6 | ABX08336_16 |
| 10 | 32.4 | 5.8 | 110000 | 6 | ABX08336_15 |
| 11 | 32 | 5.8 | 549 | 7 | ACA30202 |
| 12 | 32 | 5.8 | 936 | 4 | AAH31609 |
| 13 | 31.8 | 5.7 | 528 | 7 | AAH73424 |
| 14 | 31.8 | 5.7 | 969 | 6 | ABK65284 |
| 15 | 31.8 | 5.7 | 969 | 6 | ABK65284 |
| 16 | 31.8 | 5.7 | 969 | 6 | ABK65284 |
| 17 | 31.8 | 5.7 | 969 | 6 | ABK65284 |
| 18 | 31.8 | 5.7 | 4715 | 4 | AAH74514 |
| 19 | 31.6 | 5.7 | 409 | 4 | AAH83491 |
| 20 | 31.6 | 5.7 | 2150 | 6 | ABK85746 |
| 21 | 31.6 | 5.7 | 2217 | 9 | ADC92937 |
| 22 | 31.6 | 5.7 | 17286 | 4 | AAH36869 |
| 23 | 31.6 | 5.7 | 17286 | 7 | ABX59857 |

| | | | | | | |
|----|------|-----|--------|---|----------|--------------------|
| 24 | 31.4 | 5.7 | 2058 | 5 | AAE93936 | AAE93936 DNA encod |
| 25 | 31.2 | 5.6 | 1300 | 6 | ABK65274 | ABK65274 Arabidops |
| 26 | 31.2 | 5.6 | 1300 | 7 | ADA30808 | ADA30808 Plant yie |
| 27 | 31.2 | 5.6 | 1308 | 7 | ACA26237 | ACA26237 Prokaryot |
| 28 | 31.2 | 5.6 | 11485 | 9 | ADB79904 | ADB79904 Mouse put |
| 29 | 31.2 | 5.6 | 15275 | 4 | AAE35975 | AAE35975 Human car |
| 30 | 31.2 | 5.6 | 15275 | 9 | ADB46669 | ADB46669 Human car |
| 31 | 31 | 5.6 | 918 | 2 | AAT28561 | AAT28561 Bacterial |
| 32 | 31 | 5.6 | 918 | 2 | ABK76986 | ABK76986 Bacterial |
| 33 | 31 | 5.6 | 4590 | 5 | AAH24065 | AAH24065 Yeast AD |
| 34 | 31 | 5.6 | 5162 | 2 | AAV69919 | AAV69919 Expressio |
| 35 | 31 | 5.6 | 5162 | 2 | AAV69922 | AAV69922 Mltagenic |
| 36 | 31 | 5.6 | 5262 | 2 | AAV69927 | AAV69927 Promoter |
| 37 | 31 | 5.6 | 6850 | 2 | AAV69920 | AAV69920 Expressio |
| 38 | 31 | 5.6 | 105194 | 6 | ABK24122 | ABK24122 Bacterial |
| 39 | 30.8 | 5.5 | 33248 | 4 | ABK28537 | ABK28537 Genomic |
| 40 | 30.6 | 5.5 | 476 | 4 | AAH15696 | AAH15696 Probe #56 |
| 41 | 30.6 | 5.5 | 476 | 4 | ABK57784 | ABK57784 Human fce |
| 42 | 30.6 | 5.5 | 476 | 4 | AAH37368 | AAH37368 Probe #60 |
| 43 | 30.6 | 5.5 | 476 | 4 | ABK27146 | ABK27146 Probe #56 |
| 44 | 30.6 | 5.5 | 476 | 4 | AAK31484 | AAK31484 Human bon |
| 45 | 30.6 | 5.5 | 476 | 4 | AAK05851 | AAK05851 Human bra |

ALIGNMENTS

| | | |
|----------|---|--------------------------------|
| RESULT 1 | AA79331 | AA79331 standard; DNA; 605 BP. |
| ID | AA79331 | |
| XX | AA79331 | |
| AC | AA79331 | |
| XX | AA79331 | |
| DT | 27-AUG-2003 (revised) | |
| DT | 16-FEB-1998 (first entry) | |
| XX | | |
| DE | DNA encoding Lall.1 esterase es2. | |
| XX | | |
| KW | Esterase; thermostable enzyme; ester; chiral compound; cheese; pulp; | |
| KW | paper; lignin removal; sugar; lignocellulose; disease resistance; | |
| KW | feedstuff; ss. | |
| XX | | |
| OS | Unidentified. | |
| XX | | |
| PN | WO9730160-A1. | |
| XX | | |
| PD | 21-AUG-1997. | |
| XX | | |
| PF | 11-FEB-1997; 97WO-US002039. | |
| XX | | |
| PR | 16-FEB-1996; 96US-00602359. | |
| XX | | |
| PA | (RECO-) RECOMBINANT BIOCATALYSIS INC. | |
| XX | | |
| PI | Robertson DE, Murphy D, Reid J, Maffia AM, Link S, Swanson RV; | |
| XX | Warren PV, Koenicka A, Callen W; | |
| XX | WPI; 1997-425035/39. | |
| DR | P-PSDB; AAW23078. | |
| XX | | |
| PT | Nucleic acid encoding heat stable esterase from thermophilic bacteria - | |
| PT | which is active in organic solvents, useful in cheese or paper | |
| PT | manufacture, and to study plant resistance to disease. | |
| XX | | |
| PS | Disclosure; Page 63-64; 113pp; English. | |
| XX | | |
| CC | This DNA sequence codes for Lall.1 esterase es2 (AAW23078). Newly | |
| CC | identified polynucleotides (AA79331-40) encoding esterases (AAW23069- | |
| CC | 88), some of which are claimed, can be used for recombinant production of | |
| CC | the enzymes in host cells, and as probes to identify related sequences. | |
| CC | The esterases are stable at high temperature and in organic solvents. | |
| CC | Making them superior for use in production of pure chiral compounds used | |
| CC | in pharmaceutical, agricultural and other chemical industries. A method | |

CC is claimed for transferring an amino group from an amino acid to an alpha-keto acid using a claimed esterase. The enzymes may also be useful as ripening starters in cheese making, in lignin removal in paper and pulp manufacture, in carbohydrate derivative synthesis, in fermentable sugar production from lignocellulosic waste, in the study of plant wall structure, plant resistance to disease and organic matter decomposition and to select plants bred for production of highly degradable animal feeds. (Updated on 27-AUG-2003 to correct OS field.)

5Q Sequence 605 BP; 173 A; 115 C; 135 G; 182 T; 0 U; 0 Other;

| | | | | | | | |
|-----------------------|-------|--------------|-----|------------|-----|--------|-----|
| Query Match | 76 | Score | 76 | DB | 2 | Length | 605 |
| Best Local Similarity | 50.6% | Pred. | No. | 2.1e-15 | | | |
| Matches | 265 | Conservative | 0 | Mismatches | 250 | Indels | 9 |
| | | | | | | Gaps | 3 |

| | | |
|----|---|-----|
| QY | CAGATTATCTTGTCATGGACTCTATATGCAAGGCTGTGTATCATCCAGCTTACAT | 63 |
| Db | 13 CAGCTTATGTGTTTACATGAGCTTATATATGTCTGGCTGTGTATGCGGCCGCTTATGTTG | 72 |
| QY | 64 CGTCTGCATAAATTTGGGTTATCGTACTCAACCAATTTAGTACACTCACTCGCTATGAT | 123 |
| Db | 73 CGTCTATAGAGTCCGGGGGTTAAAGTTTAACTTAACTCAATCAATCTCGAACCTTAAT | 132 |
| QY | 124 GATGAGGCCATTTTTTGGCCGCTTGAACCATGAGTCACTAGCTCGCTCATATGCTTTA | 183 |
| Db | 133 CGAGATCTTATTTTTTATGCGCAATATGATGAGTTATTTAGCAATG--AGCTTCTGCTTTA | 185 |
| QY | 184 GTCCGACACAGTTTGGGCGGATTTGTGATCAACGTTATCTAGAAATCGCGGCACCGTCC | 243 |
| Db | 190 GTGTGTACTCTATGGGGGGCTTAGTTGTCTCGGCTATTTAGAGCCAACTCAGGCCA | 249 |
| QY | 244 TGTGAACCCGTCCCATGTCGTGCGCATCGGCTCACTTCGACAGAGGCTTCATTTGTC | 303 |
| Db | 250 AGTATCATGTTGAAAAGTATCACTTATGGAAGCCACATCTGCGAGCCATATTTGCT | 309 |
| QY | 304 AATTAAATTAGCAATTTAGGTTTAAAGGGTGGCACTAGTAATTCAGCAAAATTTGGGTTA | 363 |
| Db | 310 GAAAAATTCAGCAAAAAGGGTTTCGAGCTATTATTTAAAAATAGCGTTGAGTTT---TTA | 366 |
| QY | 364 AAAGAAACAGACAGCAAGATCCGCGTATCCACAAAATCGGAGATTTGCGAGAACGATA | 422 |
| Db | 367 CTCTCTAAGATGTGATTTGGCTTTTAAAGCCAGCTATATAGCATTTGCCGGGACCTTA | 428 |
| QY | 424 CCTTTAGGGCTGGCAGCCCTTTTACTCGCGATCCATGGAAGCTCGATGTATCCGTCA | 483 |
| Db | 427 CCGATTTGGCTTAATGCACTCATTTGTAATAAGGACGCC---GCTGTAGTGGCATGTATTG | 483 |
| QY | 484 GTAGAAGAAACAAAATATGCTGCGATGACAGCATCATATGCGAAT | 527 |
| Db | 484 CTAGATGAACCAAGCTTAAGGGTATGGCTGAACACAGAGTATT | 527 |

RESULT 2

ID ADA71938 standard; DNA; 2000 BP.

AC ADA71938;

DT 20-NOV-2003 (first entry)

DE Rice gene, SEQ ID 5263.

KW Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.

05 *Oryza sativa*.

PN WO2003000898-A1

PD 03-JAN-2003.

PF 22-JUN-2001; 2001WO-IB001105.

PR 22-JUN-2001; 2001WO-IB001105
XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Chang H, Chen W, Cooper B, Glazebrook J, Gott SA, Hou Y,
Wang X, Zhang Y, Zhao Z, Zhu T

XX
23
EFT 0000 125000/17

XX
The 14th

pathogenic infection for conferring resistance or tolerance to a plant to

PT gene expression.

PS Claim 27; SEQ ID NO 5263; 899pp; English.

CC The present inve

CC comprises identifying a gene whose expression

the inhibitory interaction of plant gene expression represses the expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M1) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

| | | | | |
|-----------------------|-------|-------------------|-----------------|--------------|
| Query Match | 6.7%; | Score 37.2; | DB 7; | Length 2000; |
| Best Local Similarity | 7.7%; | Pred. No. 0.11; | | |
| Matches | 39; | Conservative 231; | Mismatches 234; | Indels 0; |
| | | | | Gaps 0 |

Oy 30 TATCAGGCTGGTGAATGATCCGCTTAGCATCGCTGCATATAATGGGTATTCGAC 89
 Db 523 TMRCAKRTYSASAARARCMYRBRKGYTMAMMMKRXRMRYMYDMMMYKRYKSKSMYIC 582
 Oy 90 TCAACCAATTAGCTACAACTCACTCACTGCTATCGATGATGAGCCATTTTTCGCCGCTTTGA 149
 Db 583 KMSYTAACMSARARAGAKMCRKSSMSMSKMSBSRCKRKCASRSSAKRYAMMGMTSG 642
 Oy 150 CCGATGCGTCACTATGACGCTGCGCTTANCGCTTAGCGAGACAGTTTGGCGGATGGT 209
 Db 643 SRMSRWKSYTYTWRKMSKMSKSTCTWMTYMSKTYTAKGSIYRYRPMCMYMMMYRY 702
 Oy 210 GATCAAAAGTTATAGAAATGCGCGGACCGCTCGTGAAAACCTCTCCCATGTGCTGC 269
 Db 703 RSYMTYAMWYTSRPMAMTGKYSGRWTSWYKCKMSKRSWMYYSMMWMAATWKKMR 762
 Oy 270 CATCGGCTCACTTGCAGAAGAGCTTCATCTGCAATAAATAGCAATTAGGTTGAG 329
 Db 763 RYATPRMMMMYRYSMKYTWCTMGYMMYMYRTYMKRIMYKCTKTYIYWSATYWTGT 822
 Oy 330 GGTGCACTAGATBATTACGACGAAATTTGGTTAAAGAACGACGAGATCCCGCTGA 389
 Db 823 AAMWMAKTYOMMGMTGAKTRBARAKARYMMKMAIWCATKRWMTGKGKAWMTWMAKMR 882
 Oy 390 TCCCAAAAATCAGGCAAGTATTCGACGAAGAACATACCTTTAGGCGTGGCAACCTTTACT 449
 Db 883 KYMSWMAWYXXKTRTRYKTTWMAKRWMSMAIYMRXMYMGASAMMMWCGGRNGWTKY 942
 Oy 450 GCGGATCCATGACCTCCGATGGTATGCGCTACAGTGAAGAAAACAAAATAGCTGCAT 509
 Db 943 WYWTCTTHKACGBAITKXMCAGMMWAMYSWTHTYMRITWRMMWASRTAKMRPMMWK 1002
 Oy 510 GACGAGATATATCGCGAATTCAC 533
 Db 1003 TRAWSKYARAIFYKAGCACTTAC 1026

RESULT 3

ID AAS46747 standard; DNA; 26997 BP.

XX AAC46747;
AC
XX
DT 18-DEC-2001 (first entry)
DE Tumour suppressor gene derived chemically modified sequence #471.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KM tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002955.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX MPI; 2001-602752/68.
XX
XX
XX Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX
XX
XX Claim 1; SEQ ID NO 471; 27pp; English.
XX
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA), e.g. with
XX bisulphite, of genes associated with tumour suppression and oncogenes
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX 500 are missing from the sequence listing) sequences (Ss) and sequences
XX complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX probes for detecting the cytosine methylation state and/or single
XX nucleotide polymorphisms and also to be used in an array for analysing
XX diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX probes can also be used in a method for ascertaining genetic and/or
XX epigenetic parameters for the diagnosis and/or therapy of existing
XX diseases or the predisposition to specific diseases, by analysing
XX cytosine methylations. The parameters may be compared to another set of
XX genetic and/or epigenetic parameters, the differences serving as basis
XX for diagnosis and/or prognosis events which are disadvantageous to
XX patients. The present sequence is one of the 533 genomic sequences
XX derived from tumour suppressor genes and oncogenes. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 26997 BP; 7233 A; 354 C; 6190 G; 13220 T; 0 U; 0 Other;
SQ

Query Match 6.6%; Score 36.8; DB 4; Length 26997;
Best Local Similarity 47.1%; Pred. No. 0.57;
Matches 113; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 277 TCACCTTGCAAGAGGCTTCATTCATATAAATTGACATTTAGGTTAGGGGTGCA 336
DB 13416 TCTCATGAAAGTAACTAACATCCATATACAAATTAACCTAAATACGTTAATTC 13357
QY 337 CTAGGTAAATTCAGCAAAATTGGGTTAAAGAACGACGACGAATCCCGCTATCCACA 396
DB 13356 CTACCAACATTTATCAACTTCGACACTAATACCACTCGAACCACCAATCTTTAAACAA 13297

QY 397 AAATCAGGCGATATTCAGAGACATACCTTTAGGGCTCGCAGCCTTTACTGCCGAT 456
DB 13296 ATAAATATAAAAACTACAAAAACGCTACCTTATATCTACAAATATCCCGACTCTAC 13237
QY 457 CCACTGACCTCCGATGAGTACCCGTCACAGTAGAAGAAACAAATATCGGATGACAGAT 516
DB 13236 CCACTGACCTCCGATGAGTACCCGTCACAGTAGAAGAAACAAATATCGGATGACAGAT 13177

RESULT 4
ID ACA64845 standard; DNA; 115218 BP.
XX
XX ACA64845;
XX
XX 27-JUN-2003 (first entry)
XX
XX Human HNRP GP43 DNA corresponding to AL034397.
XX
XX Human; chronic inflammatory joint disease; infection; tumour;
XX anti-inflammatory; cytostatic; antiarthritic; antineumatic;
XX immunosuppressive; gene therapy; etiological pathogenicity; ds.
XX
XX Homo sapiens.
XX
XX DE10127572-A1.
XX
XX 05-DEC-2002.
XX
XX 30-MAY-2001; 2001DE-01027572.
XX
XX 30-MAY-2001; 2001DE-01027572.
XX
XX (PAT-) PATHOARRAY GMBH.
XX
XX Haeupl T, Ungethlem U, Blass S;
XX
XX MPI; 2003-240797/24.
XX
XX
XX Reagents for diagnosis, study and therapy of chronic inflammatory joint
XX and other diseases, comprises any of many specified genes or derived
XX proteins.
XX
XX Claim 1; Page; 12pp; German.
XX
XX This invention describes a novel reagent for diagnosis, molecular
XX definition and therapy of chronic inflammatory joint diseases, and other
XX inflammatory disorders, infective or tumour diseases in humans. The
XX products of the invention have anti-inflammatory, cytostatic,
XX antiarthritic, antineumatic and immunosuppressive activity and can be
XX used for gene therapy. The reagent of the invention and any proteins and
XX antibodies derived from it, are used (i) for analysing tissue and blood
XX samples for medical diagnosis; (ii) for diagnosis and characterisation of
XX chronic joint diseases, on the basis of molecular characterisation, and
XX determining the etiological pathogenicity principle of as yet
XX uncharacterised inflammatory diseases, also monitoring progression and/or
XX treatment of disease, and optimisation of therapy and (iii) for
XX developing treatments for inflammatory diseases, particularly of joints,
XX infections and tumours. ACA64801-ACA64965 represent human polymucleotides
XX used in the method of the invention
XX
XX Sequence 115218 BP; 35618 A; 23477 C; 22286 G; 33837 T; 0 U; 0 Other;
SQ

Query Match 6.4%; Score 35.6; DB 7; Length 115218;
Best Local Similarity 52.7%; Pred. No. 3.1;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGCACATATTCCTTCTGATGACCTATATGATGAGTGTATGATCGGCTTACT 60
DB 53509 ATACACTTCTCTCTTTAGGGAATCAATTTAAGAGGTTGGTATCAATGCTAGGT 53568
QY 61 CATGCTGTCATATAATTTGGTTATCTGTAACAACCATTTAGTACAACTCGCTATC 120

DB 53569 TCTTATCAAGATAAATACTATTTCATGGGAAAGACCTTTGGTCTCTACTGAATCCCTTC 53628
 QY 121 GATGATGAGGCCATTTTTCGCCGCT 146
 DB 53629 CATGGTAAAGAGAGTGTGAGCCT 53654

RESULT 5
 AAV13834/c
 ID AAV13834 standard; cDNA; 2277 BP.
 XX
 AC AAV13834;
 DT 21-JUL-1998 (first entry)
 XX
 DE Homo sapiens ambiguity-maximised telomerase protein p105 gene.
 XX
 KM telomerase; p105; treatment; prevention; cancer; restenosis;
 KM inflammation; myocardial infarction; glomerulonephritis; transplant;
 KM rejection; infection; HIV; human immunodeficiency virus;
 KM bone marrow transplants; proliferation-restricted cells;
 KM ambiguity-maximised; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT 1.2277
 FT CDS /*tag= a
 FT /product= "telomerase protein p105"
 XX
 PN WO9801543-A1.
 XX
 PD 15-JAN-1998.
 XX
 PF 08-JUL-1997; 97WO-US012297.
 XX
 PR 08-JUL-1996; 96US-00676967.
 XX
 PA (TULAR) TULARIK INC.
 PI Cao Z;
 PI WPI; 1998-101044/09.
 DR P-PSDB; AAW41927.
 XX
 PT New nucleic acid encoding human telomerase protein p105 or its fragments
 PT - used for therapeutic modulation of telomerase activity and for
 PT screening for potential modulators of telomerase-target binding.
 XX
 PS Disclosure; Page 20-21; 32pp; English.
 XX

The sequence is that of an ambiguity-maximised human telomerase protein coding sequence. The sequence, or specific fragments of it, can be used to modulate expression of a telomerase transcript (by hybridising to it intracellularly), e.g. for treatment or prevention of cancer, restenosis, inflammation, myocardial infarction, glomerulonephritis, transplant rejection and infections (e.g. with human immunodeficiency virus). It can be used to express recombinant telomerase protein which can be used to screen for agents, e.g. antibodies, that modulate binding of human telomerase to its binding target. Those that inhibit telomerase activity can be used to treat the conditions listed above, while those that are agonists can be used to extend the life of proliferation-restricted cells, especially normal somatic cells, e.g. in cases of hypersensitivity or atrophy, also to improve production of recombinant proteins by maximising cell density and survival and expansion of precursor cells being used for bone marrow transplants. They may also be used for diagnosis. Other uses of telomerase proteins are isolation, enrichment and concentration of telomerase RNA or proteins; as immunogens; in cell therapy; as reagent where nascent oligonucleotides of known structure are needed (e.g. for tagging native nucleic acid molecules) and for regulating cell growth/density tolerance. The agents and the telomerase proteins should be very specific, e.g. they are selective for cancer

CC cells without harming somatic cells
 XX
 SQ Sequence 2277 BP; 513 A; 212 C; 395 G; 217 T; 0 U; 940 Other;
 Query Match 6.2%; Score 34.4; DB 2; Length 2277;
 Best Local Similarity 34.5%; Pred. No. 1.1;
 Matches 51; Conservative 25; Mismatches 72; Indels 0; Gaps 0;

QY 11 TCTGTTCATGACCTCTATATGATGCTGTGTAATGATCCGCTATGATCTCTCC 70
 DB 662 TTTTYYTNACNSWYTCYGRITGNTNSWYTCGNSWYTTTCYTCNCDDATNGNSWN 603
 QY 71 ATAAATGGGTATCGACTCAACACATAGCTACACTCACTCGCTATGATGATGAG3 130
 DB 602 ACNSWITGATGTCCTTTRATYTRTCYTTNGCNAAGCCCATCNAACNGMANGTNGX 543
 QY 131 CCAATTTTCGCCGCTTGACCGATGCT 158
 DB 542 CCYTDAATYTCYTTCAATRTCAATCCYT 515

RESULT 6
 AAV05370/c
 ID AAV05370 standard; cDNA; 2277 BP.
 XX
 AC AAV05370;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human telomerase p105 subunit synthetic gene.
 XX
 KM Telomerase; p105; human; cell replication; cancer; restenosis;
 KM multiple sclerosis; inflammation; rheumatoid arthritis;
 KM myocardial infarction; glomerulonephritis; transplant rejection;
 KM infection; therapy; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO9801542-A1.
 XX
 PD 15-JAN-1998.
 XX
 PF 08-JUL-1997; 97WO-US012296.
 XX
 PR 08-JUL-1996; 96US-00676974.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PI Collins K;
 PI WPI; 1998-101043/09.
 XX
 PT New nucleic acid encoding human telomerase proteins or their fragments -
 PT useful for therapeutic modulation of telomerase activity and for
 PT screening for potential modulators of telomerase-target binding.
 XX
 PS Disclosure; Page 19-20; 32pp; English.
 XX

This polynucleotide comprises a synthetic, ambiguity-maximised DNA coding for the p105 subunit (see AAW46593) of human telomerase. It is based on an isolated cDNA clone (see AAV05369) for p105 and encompasses all possible nucleic acids encoding the full-length protein. The invention provides methods relating to human telomerase and related nucleic acids, including the subunit proteins p140, p105, p48 and p43. The proteins may be produced recombinantly from transformed host cells or purified from human cells. Also included are human telomerase RNA (see AAV05373) and functional derivatives (see AAV05374 and AAV16092-93), as well as p105 synthetic DNA sequences (AAV05370-72). The invention also provides isolated telomerase hybridisation probes and primers capable of specifically hybridising with the telomerase gene, telomerase-specific binding agents such as specific antibodies, and methods of making and using the subject compositions in diagnosis (e.g. genetic hybridisation

Claim 14; SEQ ID NO 1536e; 1766pp; English.

Claim 14; SEQ ID NO 15368; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 613 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the targete prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPRO at http://wipro.int/pubmed/published_pat_sequences

Sequence 948 BP; 370 A; 114 C; 204 G; 260 T; 0 U; 0 Other;

Query Match 5.9%; Score 32.8; DB 7; Length 948;

Matches 103; Conservative 0; Mismatches 117; Indels 0; Gaps 0

| | | | |
|----|-----|---|-----|
| QY | 336 | ACAGGAAATTCACGCAAAATTTGGGTTAAAGAAACGACGACGATCCCGTATCCCA | 395 |
| Db | 561 | AATAACTGATTACACAAACTCAACATAGAAAGTATGGGAAATGGTACTATTAAGCCAGA | 622 |
| QY | 396 | AAAAATCAGCGCATTTGACGAAACGATACCTTTAGGGCTTGGCGACGCTTTACTGCGCA | 455 |
| Db | 621 | AGAGCAATATGTTGGCACAAAGATCTTATAGAGCATTTTAAATTATTATACCTCT | 680 |
| QY | 456 | TCCACTGGACCTCCGATGGTACCGTCACAGTGAAGAAACCAAAATAGCTGGCATGACGA | 515 |
| Db | 681 | TACGATCTGCGCATGATATGTGAGATTAATGGTGAAGAAAAGAAAGATTAAGAGAAATA | 740 |
| QY | 516 | TCATATGCGCATTCGACCACTTCATACGAGAAATGCTGT | 555 |
| Db | 741 | GGTCTTTGAGTAGCTATCGAAGAACTTGATCTTTAGATT | 780 |

RESULT 9

ABX08336 16/c
Cell Culture (17) of ABY08336 from Page 1600001 (Human rhosphodiesterase 4D (PDE4D

WP Sequence split into 17 fragments LOCUS ABX083336 Accession Abx083336

| | | | |
|----|-------------|---|--------|
| WP | ABX08336_00 | 1 | 110000 |
| WP | ABX08336_01 | 1 | 210000 |

| | | | |
|----|-------------|--------|--------|
| WP | ABX08336_02 | 200001 | 310000 |
| WP | ABX08336_03 | 300001 | 410000 |
| WP | ABX08336_04 | 400001 | 510000 |
| WP | ABX08336_05 | 500001 | 610000 |
| WP | ABX08336_06 | 600001 | 710000 |
| WP | ABX08336_07 | 700001 | 810000 |

| | | |
|----|-------------|-----------|
| WP | ABX08336_08 | 9.00001 |
| WP | ABX08336_09 | 900001 |
| WP | ABX08336_09 | 10.100001 |
| WP | ABX08336_10 | 1.1100001 |
| WP | ABX08336_11 | 1.000001 |
| WP | ABX08336_11 | 1.000001 |
| WP | ABX08336_12 | 1.1200001 |
| WP | ABX08336_12 | 1.200001 |
| WP | ABX08336_13 | 1.300001 |
| WP | ABX08336_13 | 1.400001 |
| WP | ABX08336_14 | 1.500001 |
| WP | ABX08336_15 | 1.600001 |
| WP | ABX08336_16 | 1.600001 |
| WP | ABX08336_16 | 16.100001 |

| | | | | |
|-------------|-------|-------------|-------|---------------|
| Query Match | 5.88; | Score 32.4; | DB 6; | Length 91080; |
|-------------|-------|-------------|-------|---------------|

| | | | | | | | | | |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|
| Matches | 81; | Conservative | 0; | Mismatches | 81; | Indels | 0; | Gaps | 0; |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|

| | | | |
|----|------|--|------|
| Qy | 274 | GGCTCACCTTGGCAAGAGCTTCATATGTCAATAAAATTGCAATTAGTTTAGGGCTG | 333 |
| Db | 2603 | GGCTCATCTTACATATTTTCCTTCCTGAGTCCCTGAAATCGCATATTTTTCCAAAG | 2544 |
| Qy | 334 | GCACTAGTAGTATTCAGCAGATTGGGTTAAAGACACGACGACGATCCCGATGCA | 393 |
| Db | 2513 | TCTTGTCCTCTTTTACAGAGAAATGTATTGAAACAAATAGAGGTCTAAGTGTGT | 248 |
| Qy | 394 | CAAAATACGACAGTATTTCAGAGAACGATACCTTTAGGGCTG | 435 |
| Db | 2493 | CATGCTACTGGGTATGTATGCTTCTTAGACCTCTCGAGCTG | 2442 |

RESULT 10

ABX08336_15/c

WP Sequence split into 17 fragments LOCUS ABX08336 Accession Abx08336

WP ABX08336_00 1

| | | | | |
|----|-------------|--------|--------|--------|
| WP | ABX08333_02 | 20001 | 31000 | 41000 |
| WP | ABX08333_03 | 30001 | 41000 | 51000 |
| WP | ABX08333_04 | 40001 | 51000 | 61000 |
| WP | ABX08333_05 | 50001 | 61000 | 71000 |
| WP | ABX08333_06 | 60001 | 71000 | 81000 |
| WP | ABX08333_07 | 70001 | 81000 | 91000 |
| WP | ABX08333_08 | 80001 | 91000 | 101000 |
| WP | ABX08333_09 | 90001 | 101000 | 111000 |
| WP | ABX08333_10 | 100001 | 111000 | 121000 |
| WP | ABX08333_11 | 110001 | 121000 | 131000 |
| WP | ABX08333_12 | 120001 | 131000 | 141000 |
| WP | ABX08333_13 | 130001 | 141000 | 151000 |
| WP | ABX08333_14 | 140001 | 151000 | 161000 |
| WP | ABX08333_15 | 150001 | 161000 | 171000 |
| WP | ABX08333_16 | 160001 | 171000 | 181000 |

| | | | | |
|-------------|-------|-------------|-------|----------------|
| Query Match | 5.8%; | Score 32.4; | DB 6; | Length 110000; |
|-------------|-------|-------------|-------|----------------|

| | | | | | | | | | |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|
| Matches | 81; | Conservative | 0; | Mismatches | 81; | Indels | 0; | Gaps | 0; |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|

QY 274 GGCTCACCTTTGGAGAGACTTCCANFTGCAATATAATATGCAATTAGTTTGGGGTG 333

Db 102603 GGCTCATCTTACATATTTCTCTCCGAGTCCCTAGAAATCAGCATATTTTTTCAAGAG 102544

QY 334 GCACATAGATTAATCAGCAGAAATTTGGGTAAAGACACGACGAGATCCCGCTATCCA 393

Db 102543 TCTTGGTCTTTTACGAGAAATGATATTAGAACCAAAATGAGGGTCTTAATGTGCT 102486

QY 394 CAAAATCAGCAGATTTGAGAGAAAGATACCTTTAGGGCTG 435

Db 102483 CATGGCTACTGGGTATGGATCTTTAGAACCTCTCTGAGCTG 102442

RESULT 11

ID ACA30202 standard; DNA; 549 BP.

ACA30202;
AC
XX

PT 19-JUN-2003 (first entry)
 XX Prokaryotic essential gene #11859.
 DE Prokaryotic essential gene #11859.
 XX Antisense; ds; prokaryotic essential gene; cell proliferation;
 KM drug design; gene.
 XX Campylobacter jejuni.
 OS
 XX WO200271183-A2.
 PN
 XX 03-OCT-2002.
 PD
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 PF
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 PI Wang L, Zamudio C, Malone C, Hasebeek R, Ohlsen KL, Zyskind JM;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR F-PSDB; AB026332.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 18072; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-regulated gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 549 BP; 198 A; 64 C; 116 G; 171 T; 0 U; 0 Other;

Query Match 5.8%; Score 32; DB 7; Length 549;
 Best Local Similarity 65.3%; Pred. No. 3.7;
 Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 332 TGGCACTAGTAATTGACGAAATTTGGTTAAAGAACGACGACGAAATCCCGTATC 391
 DB 428 TGATACTGTGAAATCTCGCAAAATTTGTGCTGTAATTAACGATGAAGATCAGCTTTC 487
 QY 392 CACAAAATCAG 403
 DB 488 TAAACATTCG 499

RESULT 12

AAH31609 standard; DNA; 936 BP.

AC AAH31609;
 DT 30-JUL-2001 (first entry)
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 182.
 XX

KM Human, olfactory receptor; OR; primary scent determination;
 KM secondary scent determination; polypeptide library; odour receptor;
 KM scent profile; scent fingerprint; scent representation; ds.
 XX

OS Homo sapiens.

PN W0200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US027582.

PR 08-OCT-1999; 99US-0158615P.

PR 24-FEB-2000; 2000US-0184809P.

PA (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Ballenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

DR WPI; 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists.
 XX

PS Claim 8; Page 248; 1857pp; English.

CC The present sequence is one of a number of isolated polynucleotides which
 CC encode polypeptides involved in olfactory sensation. The polynucleotides
 CC can be used in screening for olfactory agonists and antagonists. The
 CC methods allow for the determination of primary scents and the
 CC identification of the odour receptors used to detect these primary
 CC scents. The methods also enable determination of secondary scents and the
 CC identification of combinations of odour receptors that are involved in
 CC detecting such secondary scents. This enables the construction of a scent
 CC representation (also called a scent fingerprint or scent profile), which
 CC may be used to re-create and edit scents. Libraries of olfactory
 CC receptors are useful for determining the interaction pattern of a
 CC composition with the receptors, and can be used for determining
 CC differences in the olfactory faculties of different individuals
 XX

SQ Sequence 936 BP; 214 A; 232 C; 174 G; 316 T; 0 U; 0 Other;

Query Match 5.8%; Score 32; DB 4; Length 936;
 Best Local Similarity 62.5%; Pred. No. 4.8;
 Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 259 CATGCGTCCCATCGGCTCACCTTGCAGAGAGCTTCATTGTGCATATAAATTGACAA 318

DB 358 CGTTTATGCGATCCGCAATCTTTGAGATATGCTTCATTTCACCAATCTAGAGTC 417

QY 319 TTAGTTTAGGGGTGGCACT 338

DB 418 ATAGCGTTAGGAGTGGAGT 437

RESULT 13
ACF73424/C
ID ACF73424 standard; DNA; 528 BP.
XX
XX ACF73424;
AC
XX 20-NOV-2003 (first entry)
DT
XX Staphylococcus aureus DNA #1104.
DE
XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KM enzymatic assay; antibiotic target; gene; ds.
XX
XX Staphylococcus aureus.
OS
XX WO200294868-A2.
PN
XX 28-NOV-2002.
PD
XX 27-MAR-2002; 2002WO-1B002637.
PF
XX 27-MAR-2001; 2001GB-00007661.
PR
XX (CHIR-) CHIRON SPA.
PA
XX Masignani V, Mora M, Scarselli M,
PI
XX WPI; 2003-120786/11.
DR
XX P-PSDB; AEM71864.
XX
XX New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX
XX Claim 6; SEQ ID NO 2207; 49pp; English.
PS
XX The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus genes of the invention
XX
XX Sequence 528 BP; 212 A; 57 C; 116 G; 143 T; 0 U; 0 Other;
SQ
Query Match 5.7%; Score 31.8; DB 7; Length 528;
Best Local Similarity 59.3%; Pred. No. 4.3;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 2 TGGAGATTTCCTGGTTCATGACCTATATGATGCGTGGTATGATCGCTTATGC 61
DB 480 TTTATCTTATCTCTTCATATGATATCCGATGATGATTTCTTCTCACTGCTC 421
QY 62 ATCGTCTGATTAATGGCTTATCGTACTCA 92
DB 420 ATCTTCTGATCATTTAGTTCTTCTTGATCA 390
RESULT 14
ABK65284
ID ABK65284 standard; cDNA; 969 BP.
XX
XX ABK65284;
AC
XX 02-JUL-2002 (first entry)
DT
XX Arabidopsis cDNA encoding a transcription factor #136.
DE

XX
XX Plant; ss; gene; transcription factor; transgenic; agriculture;
KM metabolic chemical; environmental stress; drought;
KM microbial disease resistance; herbicide resistance; seed yield;
XX fruit yield; growth rate; leaf senescence; flower senescence.
XX
XX Arabidopsis thaliana.
OS
XX WO200215675-A1.
PN
XX 28-FEB-2002.
PD
XX 22-AUG-2001; 2001WO-US026189.
PF
XX 22-AUG-2000; 2000US-0227439P.
PR 16-NOV-2000; 2000US-00713994.
PR 18-APR-2001; 2001US-00837944.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (PILG/) PILGRIM M.
PA (CREE/) CREELMAN R.
PA (DUBE/) DUBELL A J.
PA (HEAR/) HEARD J.
PA (JIAN/) JIANG C.
PA (KEDD/) KEDDIE J.
PA (ADAM/) ADAM L.
PA (RATC/) RATCLIFF O.
PA (REUB/) REUBER J L.
PA (RIEC/) RIECHMANN J L.
PA (YUGG/) YU G.
PA (PINE/) PINEDA O.
XX
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J,
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX
XX WPI; 2002-292022/33.
DR
XX P-PSDB; AAU93098.
DR
XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant.
XX
XX Claim 4; Page 558-559; 941pp; English.
PS
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
CC encoding an Arabidopsis thaliana transcription factor, their variants,
CC complementary fragments, or related polynucleotide with 318 to 958
CC sequence identity, where the plant possesses an altered trait as compared
CC to a wild-type or reference plant, or the plant exhibits an altered
CC phenotype as compared to a wild-type or reference plant, or the plant
CC exhibits ectopic expression or altered expression of one or more genes
CC associated with a plant trait as compared to a wild plant. Also included
CC are a transgenic plant comprising the polynucleotides, a computer
CC readable medium having stored sequence information, and identifying a
CC homologous sequence from a database comprising a plurality of known plant
CC sequences comprising inputting sequence information selected from one of
CC 466 fully defined sequences given in the specification. The isolated or
CC recombinant polynucleotide is used for producing a plant having a
CC modified trait, the method comprising selecting a polynucleotide that
CC encodes a polypeptide or an antisense nucleic acid, inserting the
CC polynucleotide or antisense nucleic acid into an expression vector,
CC introducing the vector into a plant or a cell of a plant to overexpress
CC the polypeptide or antisense nucleic acid, thereby producing a modified
CC plant, and selecting for a modified trait (e.g. increased production of
CC agriculturally useful proteins or metabolic chemicals, pest tolerance,
CC environmental stress response (e.g. drought), microbial disease
CC resistance, herbicide resistance, seed and fruit yield, growth rate, leaf
CC senescence and many other traits listed in the specification).
CC The present sequence is one of the 232 polynucleotides encoding an A.
CC thaliana transcription factor
XX
XX Sequence 969 BP; 267 A; 247 C; 212 G; 243 T; 0 U; 0 Other;
SQ
Query Match 5.7%; Score 31.8; DB 6; Length 969;
DE

Best Local Similarity 54.8%; Pred. No. 5.8;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Dy 89 CTCAACCATTTAGCTACAACCTACTGCGTATGATGATGAGGCATTTCGGCCGCCCTTG 148
Db 392 CTCAAACCAACCAACAACCTCCTCTGTCTCCTGAAAAGCAGGTGTCACGACGA 451

Qy 149 ACCGATCGCTCACTCATGCGCTCGCCCTAATGCTTTAGTCGACACACGTTTGGGCGG 203
Db 452 AACGCCCTCGTCCACTGTGCCAGGATAAGATAGAGTCAAAGACATGTGTGCGG 506

RESULT 15
ABZ13985
ID ABZ13985 standard; DNA; 969 BP.
XY

Db 452 AACGCCCTCGTCCACCTGTCCAGGATAAGATAGAGTCAAAGACAATGTGCGG 506

Search completed: April 30, 2004, 11:04:24
Job time : 233.167 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:27 ; Search time 1575.49 Seconds
(without alignments)
10519.601 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagatattctctgttca.....cttcacagagagcgcgtc 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est4:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: gb_est1:*
29: gb_est2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| 1 | 35.2 | 6.5 | 526 | 12 | BG924185 |
| 2 | 35.6 | 6.4 | 583 | 12 | BG776621 |
| 3 | 35.6 | 6.4 | 704 | 10 | BR211558 |
| 4 | 35.6 | 6.4 | 715 | 12 | BG493794 |

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| 5 | 35.6 | 6.4 | 816 | 12 | BG497402 |
| 6 | 35.2 | 6.3 | 498 | 28 | AQ451880 |
| 7 | 35.2 | 6.3 | 1201 | 13 | BX381961 |
| 8 | 35 | 6.3 | 942 | 29 | CG098039 |
| 9 | 34.6 | 6.2 | 931 | 29 | CG711514 |
| 10 | 34.2 | 6.2 | 742 | 14 | CAB08007 |
| 11 | 33.8 | 6.1 | 434 | 13 | BQ494475 |
| 12 | 33.8 | 6.1 | 433 | 13 | BQ493352 |
| 13 | 33.8 | 6.1 | 485 | 13 | BQ499441 |
| 14 | 33.8 | 6.1 | 593 | 13 | BE776398 |
| 15 | 33.8 | 6.1 | 624 | 13 | BX311247 |
| 16 | 33.8 | 6.1 | 770 | 14 | CP408002 |
| 17 | 33.8 | 6.1 | 787 | 29 | CNS01087 |
| 18 | 33.8 | 6.1 | 797 | 29 | CC554553 |
| 19 | 33.6 | 6.1 | 435 | 13 | BX34834 |
| 20 | 33.6 | 6.1 | 485 | 13 | BQ493087 |
| 21 | 33.6 | 6.1 | 575 | 10 | AW42644 |
| 22 | 33.6 | 6.1 | 606 | 13 | BQ262340 |
| 23 | 33.6 | 6.1 | 661 | 12 | BG302675 |
| 24 | 33.6 | 6.1 | 1223 | 12 | BT952284 |
| 25 | 33.4 | 6.0 | 1456 | 12 | BX376686 |
| 26 | 33.4 | 6.0 | 1456 | 12 | BT957625 |
| 27 | 33.2 | 6.0 | 228 | 9 | AT210254 |
| 28 | 33.2 | 6.0 | 540 | 10 | AW116871 |
| 29 | 33.2 | 6.0 | 628 | 12 | BM332415 |
| 30 | 33.2 | 6.0 | 766 | 28 | BZ193724 |
| 31 | 33.2 | 6.0 | 765 | 14 | CAB09511 |
| 32 | 33.2 | 6.0 | 845 | 14 | CAB09398 |
| 33 | 33.2 | 6.0 | 1068 | 29 | CNS037RT |
| 34 | 33 | 5.9 | 247 | 9 | AT209463 |
| 35 | 33 | 5.9 | 288 | 9 | AT209943 |
| 36 | 33 | 5.9 | 396 | 13 | BY582873 |
| 37 | 33 | 5.9 | 441 | 28 | AQ110347 |
| 38 | 33 | 5.9 | 733 | 33 | CC172630 |
| 39 | 33 | 5.9 | 977 | 29 | CG684361 |
| 40 | 32.8 | 5.9 | 722 | 28 | BH715417 |
| 41 | 32.8 | 5.9 | 722 | 28 | BH533401 |
| 42 | 32.8 | 5.9 | 1047 | 29 | CNS026EJ |
| 43 | 32.8 | 5.9 | 1101 | 29 | CNS002EJ |
| 44 | 32.6 | 5.9 | 495 | 13 | BQ107442 |
| 45 | 32.6 | 5.9 | 495 | 14 | CF665213 |

ALIGNMENTS

RESULT 1
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT

BG924185 526 bp. mRNA. linear. EST 05-JUN-2001
Oo_ad_03B06 LambdaGAT11FOR Osteragia osteragi adults Osteragia
Osteragi CDNA clone Oo_ad_03B06 5' similar to pIR171628
Propionyl-CoA carboxylase (EC 6.4.1.3) Beta chain f5284.1. mRNA
Sequence.
BG924185
BG924185.1 GI:14317857
EST
Osteragia osteragi
Osteragia osteragi
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongyloidea;
Trichostrongyloidea; Haemonchidae; Osteragiinae; Osteragia.
1 (bases 1 to 526)
Blaxter M.L., Parkinson J., Whitton C., Daub J., Gulliano D.,
Hall N., Quayle M. and Bartell B.
Edinburgh University/Sanger Centre Nematode EST Project
Unpublished (2000)
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: malk.blaxter@ed.ac.uk

The library was prepared by Isabel Vercauteren for Prof J Verduyse, Ghent University, Belgium. Sequencing was performed by Claire Whitton ICAPB, University of Edinburgh

PCR Primers
FORWARD: Lambdag11ROR
PLACE: 03 row: E column: 06
Seq primer: Lambdag11ROR
High quality sequence stop: 438.
Location/Qualifiers

1..526

/organism="Ostertagia ostertagi"

/mol_type="mRNA"

/db_xref="taxon:6317"

/clone="Oo_ad_03E06"

/sex="mixed"

/dev_stages="adult"

/clone_lib="Ostertagia ostertagi adults"

/note="Vector: Lambda g11; Site 1: EcoRI (5'end); Site 2: NotI (3'end); Ostertagia ostertagi is a parasite of cattle. The library was constructed from mRNA from Ostertagia ostertagi adults and produced with random hexamer primers."

ORIGIN

Query Match 6.5%; Score 36.2; DB 12; Length 526;
Best Local Similarity 60.8%; Pred. No. 3.7;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 302 TCATTAATTAATGACATTAAGTTAGGGGTGACATAGTAATTCAGCAGAAATTTGGGT 361
DB 237 TCCGTACGATTAACGATTAAGGATGGGGAGAACTCGGCAATTTTGTAGAGAAATAGGA 296
QY 362 TAAAGAAACACGACGACGATCCGCTATCCACAAA 398
DB 297 TAAATGAGAAACGACGATCAGAACTACACATTA 333

RESULT 2

BG776621/c 583 bp mRNA linear EST 15-MAY-2001
LOCUS 602663771F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4809040 5',
DEFINITION mRNA sequence.
ACCESSION BG776621
VERSION BG776621.1 GI:14046938
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 583)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM62 row: n column: 17
High quality sequence stop: 583.
Location/Qualifiers

FEATURES

source

1..583
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4809040"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (T1 phage-resistant)"

ORIGIN

Query Match 6.4%; Score 35.6; DB 12; Length 583;
Best Local Similarity 52.7%; Pred. No. 6;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGACATTAATCTTGTGATGAGACTCTATATGATGAGCTTGATATGATCGCTTACT 60
DB 500 ATACACTTTTCTCTTTGAGGAATCAATTAAGAGTTGTGTAACATGATGAGT 441
QY 61 CATGCTGATTAATGAGTTATGCTACTCAACCATTAAGTCAACTCACTCGTATC 120
DB 440 TCTATCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 381
QY 121 GATGATGAGCCATTTTCCGCCCT 146
DB 380 CATGTTAAGAGATGTTGAGCCT 355

RESULT 3

BP211558/c 704 bp mRNA linear EST 06-NOV-2000
LOCUS 601812246F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4046747 5',
DEFINITION mRNA sequence.
ACCESSION BP211558
VERSION BP211558.1 GI:11105144
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 704)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM674 row: 1 column: 12
High quality sequence stop: 658.
Location/Qualifiers

FEATURES

source

1..704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4046747"
/tissue_type="from chronic myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_54"
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggcgccctcgcc); Site 2: SfiI
(ggcattatggcc); Double-stranded cDNA was prepared from
cell line RNA. 5' and 3' adaptors were used in cloning as
follows: 5'-CACGGCATTAAGCC-3' and

3' adaptor sequence:
5'-ATTCTAGAGCCGAGCCGCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

ORIGIN

Query Match 6.4%; Score 35.6; DB 10; Length 704;
Best Local Similarity 52.7%; Pred. No. 6.6;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGCAGATTATCTGTTTCATGACCTATATGACGCTGTTGATGATCGCTTACT 60
DB 502 ATACACTTCTCTCTTTTGGAGAAATCAATTAGAGAGTTGGTACCAATGCTAGCT 443
QY 61 CATGCTGTCATTAATGGGTTATGCTCAACACATTAGCTACACTGCTATC 120
DB 442 TCTTATCAAGATAAATACTATTTCATGGAAAGCCTTGGTCTACTGAATCCCTTC 383
QY 121 GATGATGAGCCATTCTTGGCCGCT 146
DB 382 CATGTAAAGAAAGAGTGTGAGCCT 357

RESULT 4

BG493794/c 715 bp mRNA linear EST 27-MAR-2001

LOCUS 602542056F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4673149 5',
DEFINITION mRNA sequence.

ACCESSION BG493794
VERSION BG493794.1 GI:13455308

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

Plate: LLCM1490 row: P column: 14

High quality sequence stop: 715.

Location/Qualifiers

1. 715

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4673149"

/tissue_type="mucoepidermoid carcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/clone_1lb="NIH_MGC_59"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:

Site1 (ggcgcctcgcc); Site_2: Site1 (ggcgcctcgcc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGACC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGCCGCGACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

ORIGIN Library."

Query Match 6.4%; Score 35.6; DB 12; Length 715;
Best Local Similarity 52.7%; Pred. No. 6.7;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGCAGATTATCTGTTTCATGACCTATATGACGCTGTTGATGATCGCTTACT 60
DB 446 ATACACTTCTCTCTTTTGGAGAAATCAATTAGAGAGTTGGTACCAATGCTAGCT 387
QY 61 CATGCTGTCATTAATGGGTTATGCTCAACACATTAGCTACACTGCTATC 120
DB 386 TCTTATCAAGATAAATACTATTTCATGGAAAGCCTTGGTCTACTGAATCCCTTC 327
QY 121 GATGATGAGCCATTCTTGGCCGCT 146
DB 326 CATGTAAAGAAAGAGTGTGAGCCT 301

RESULT 5

BG497402/c 816 bp mRNA linear EST 27-MAR-2001

LOCUS 602538689F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4659742 5',
DEFINITION mRNA sequence.

ACCESSION BG497402
VERSION BG497402.1 GI:13458919

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

Plate: LLCM1456 row: A column: 23

High quality sequence stop: 715.

Location/Qualifiers

1. 816

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4659742"

/tissue_type="mucoepidermoid carcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/clone_1lb="NIH_MGC_59"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:

Site1 (ggcgcctcgcc); Site_2: Site1 (ggcgcctcgcc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGACC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGCCGCGACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH-MGC

Library."

ORIGIN

Query Match 6.4%; Score 35.6; DB 12; Length 816;
Best Local Similarity 52.7%; Pred. No. 7.1;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATCCAGATTATTTCTTTATGATGACCTCTATAGATGCTGTGAAGATCCGCTTAG 60
Db 448 ATACACTTTCTTCCTTTTGGGGAATCAATTTAAAGAGTTGTATCAATATGTCTAGGT 389
QY 61 CACGCTGCTGAATTAATTGGGTTATGTAATCAACCATTTAGCTCAACTACCTGCGATC 120
Db 388 TCTTATCAAAATTAATACTATTTCATGGGAAAACCTTTGTCCTTACTGAATCCCTTC 329
QY 121 GATGATGAGGCATTTTTCGCCGCCCT 146
Db 328 CATGCTAAAGAAAGATTGGTAGACCT 303

| | | | | | |
|------------|---|--|-----------|--------|-----------------|
| RESULT 6 | AA451880 | 498 bp | DNA | linear | GSS 21-APR-1999 |
| LOCUS | AA451880 | | | | |
| DEFINITION | AA451880 | HS.514.21.A05 SP6E RPCR-11 Human Male BAC Library Homo sapiens genomic clone Plate=760 Col=9 Row=B, genomic survey sequence. | | | |
| ACCESSION | AA451880 | | | | |
| VERSION | AA451880.1 | GI:4593050 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 498) | | | | |
| AUTHORS | Mallatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L. | | | | |
| TITLE | Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome | | | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. | 96 (17), | 9739-9744 | (1999) | |
| MEDLINE | 99380389 | | | | |
| PUBMED | 10449764 | | | | |
| COMMENT | Contact: Mallatras GG, Wallace JC, Hood L | | | | |

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPc1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (inforesgen.com). BAC end Web Server: <http://www.hesc.washington.edu>
plate: 760 row: B column: 9
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 498.
Location/Qualifiers
1..498

```

FEATURES
source
location/Qualifiers
1..498
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=750 Col=9 Row=B"
/sex="male"
/clone_1b="RPC1-11 Human Male BAC Library"
/notes="vector: pBACs.6; Site 1: EcoRI; Site 2: EcoRI;"
Male plasmid DNA was isolated from one randomly chosen d
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACs.6 vector at EcoRI sites"

```

| | Query Match | 6.3% | Score 35.2 | DB 28 | Length 498 |
|-----------------------|--|------|---------------|-------|------------|
| Best Local Similarity | 58.7% | | Pred. No. 7.6 | | |
| Matches | 61 | | Mismatches | 43 | Indels 0 |
| | | | | | Gaps 0 |
| QY | 11 TTCTCTCATGAGACTCTATATGCATGGTTGGTATGATCCGCTTACATCGTCGC | 70 | | | |

Db 96 TTGTGTCGCTGGACCTTAGGCATGCTAGATATATATTGGCTATCCAAATGTTATG 155

Qy 71 ATAAATTGGGTTATGTACTCAACCATTAGTCAACTCACTC 114

Db 156 ATAAATTACTCATCTTTCTCTAAGAAATTGGATAGTTTAACTC 189

| | |
|------------|--|
| RESULT | 7 |
| BX381961 | |
| LOCUS | |
| DEFINITION | BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED homo sapiens cDNA clone CSDD1072YR05 3-PRIME, mRNA sequence. |
| VERSION | BX381961 |
| KEYWORDS | BX381961.1 GI:30453007 |
| SOURCE | EST. |
| ORGANISM | Homo sapiens (human) |
| | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| REFERENCE | 1. (bases 1 to 1201) |
| AUTHORS | Li,W.B., Gruber,C., Jessee,J. and Polayes,D. |
| TITLE | Full-length cDNA libraries and normalization |
| JOURNAL | unpublished (2001) |
| COMMENT | Contact: Genoscope |

FEATURES

Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - France
Email: seguet@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact: Feng Liang Email: fliang@litech.com
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS001072C030NP1.
Location/Qualifiers

```

FEATURES
    source
        location/Qualifiers
            1. 1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSOD1072YF05"
                /class_type="PLACENTA COT 25-NORMALIZED"
                /clone_1fb="Homo sapiens PLACENTA COT 25-NORMALIZED"
                /note="1st strand cDNA was primed with a NotI-oligo (dT)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoR V
                sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
    Query Match
        Query Similarity 4.9%; Pred No. 11;
        Match 23; Conservative 184; Mismatches 264; Indels 0; Gaps 0

```

[illegible]

| | | | | |
|-----------------------|--|------|--|---------------------------------------|
| Oy | | 373 | GACGACGAATCCCGCTATTCACAATAATCAGGCAGTATTTCGACGAACGATACCTTTAGG | 432 |
| Dd | | 975 | KCMCVKXVMCMCBKMKMCKKKMKYVCKKMMCMAMMMCMKMYMBAMMMBAMVYM | 1035 |
| Oy | | 433 | CTGGCGAGCCTTTTACTCGCGCATCCCACTGACCTCGATCGATCGGTACCA | 483 |
| Dd | | 1035 | MMAAAMMMCMCKMYMKKKMKKKMKMKMMMMNKVKCKMBMKSKSGCM | 1085 |
| RESULT 8 | | | | |
| CG098039 | | | 942 bp | DNA |
| LOCUS | | | PURFY34TD ZM 0.6_1.0 KB | Zea mays genomic clone ZMMBT0693F19, |
| DEFINITION | | | | genomic survey sequence. |
| ACCESSION | | | CG098039 | |
| VERSION | | | CG098039.1 | GI:33980333 |
| KEYWORDS | | | GSS. | |
| SOURCE | | | | |
| ORGANISM | | | Zea mays | |
| | | | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | |
| | | | Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD | |
| | | | clade; Panicoidae; Andropogoneae; Zea. | |
| | | | 1 (bases 1 to 942) | |
| REFERENCE | | | Whitelaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T., | |
| AUTHORS | | | Rennett,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,D. and | |
| | | | Bernstzen,J. | |
| TITLE | | | Maize Genomic Consortium | |
| JOURNAL | | | Unpublished (2003) | |
| COMMENT | | | Contact: Cathy Whitelaw | |
| | | | TIGR | |
| | | | 9712 Medical Center Drive, Rockville, MD 20850, USA | |
| | | | Tel: 301-838-5843 | |
| | | | Fax: 301-838-0208 | |
| | | | Email: whitelaw@igr.org | |
| | | | Seq primer: TP | |
| | | | Class: sheared ends. | |
| FEATURES | | | | |
| source | | | Location/Qualifiers | |
| | | | 1..942 | |
| | | | /organism="Zea mays" | |
| | | | /mol_type="genomic DNA" | |
| | | | /strain="B73" | |
| | | | /db_xref="taxon:4577" | |
| | | | /clone="ZMMBT0693F19" | |
| | | | /clone_lib="ZM_0.6_1.0 KB" | |
| | | | /note="Vector: PCR4-TOP0; Site_1: EcoRI; 0.6-1.0 kb high | |
| | | | COT selected genomic DNA library" | |
| ORIGIN | | | | |
| Query Match | | | 6.3%; Score 35; DB 29; Length 942; | |
| Best Local Similarity | | | 52.4%; Pred No. 12; | |
| Matches | | | 77; Conservative 0; Mismatches 70; Indels 0; Gaps 0; | |
| Oy | | 294 | TTCCATTGTCAATAAAATTAATGACAAATTAAGTTTAGGGGTGCACATGAGTAATTCACAGA | 353 |
| Dd | | 408 | TTATATCAGATGTATACCTCAGCAAAGTCTTTACGATTGCCAAAAGGCAATTTACGAGT | 467 |
| Oy | | 354 | ATTTCGGTTAAAAAGAACACGACGACGATCCCGCTATTCACAAAAATAGACGATATGC | 413 |
| Dd | | 468 | GTCGGGGCACATGACCAAGAAAGGGAATCAGCTGTCCAATCTCTCAGAAAAGATGT | 527 |
| Oy | | 414 | AGGAACGATACCTTTAGGGGTGCGCAG | 440 |
| Dd | | 528 | TGGCGCGAGATCTTTTAGATCATGATAG | 554 |
| RESULT 9 | | | | |
| CC711514 | | | 931 bp | DNA |
| LOCUS | | | OGFER4STC ZM 0.7_1.5 KB | Zea mays genomic clone ZMMBMA0501G18, |
| DEFINITION | | | | genomic survey sequence. |
| ACCESSION | | | CC711514 | |
| VERSION | | | CC711514.1 | GI:32116290 |
| KEYWORDS | | | GSS. | |

| SOURCE | ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--|-----------------|---|-----------------|-----------------|---------|---------|
| Zea mays | Zea mays | | | | | |
| Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea. | | | | | | |
| 1 (bases 1 to 931) | | | | | | |
| WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T., Renwick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N. | | | | | | |
| Consortium for Maize Genomics | | | | | | |
| Unpublished (2002) | | | | | | |
| Contact: Cathy WhiteLaw | | | | | | |
| TIGR | | | | | | |
| 9712 Medical Center Drive, Rockville, MD 20850, USA | | | | | | |
| Tel: 301-838-5843 | | | | | | |
| Fax: 301-838-0208 | | | | | | |
| Email: whiteLaw@tigr.org | | | | | | |
| Seq primer: TF | | | | | | |
| Class: sheared ends. | | | | | | |
| Location/Qualifiers | | | | | | |
| 1..931 | | | | | | |
| /organism="Zea mays" | | | | | | |
| /mol_type="genomic DNA" | | | | | | |
| /strain="B73" | | | | | | |
| /db_xref="taxon:4577" | | | | | | |
| /clone="ZMMBMA0501G18" | | | | | | |
| /clone_1lb="ZM 0.7-1.5 KB" | | | | | | |
| /note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb | | | | | | |
| methylation filtered genomic DNA library" | | | | | | |
| ORIGIN | | | | | | |
| Query Match | 6.2%; | Score 34.6; | DB 29; | Length 931; | | |
| Best Local Similarity | 47.5%; | Pred. No. 16; | | | | |
| Matches 103; | Conservative | 0; | Mismatches 114; | Indels 0; | Gaps 0; | |
| QY | 202 | GGATTGGTGAATCAACGTTATCTAGATTCGGCGCCACCGTCCTGAAACCCCTCCCAT | 261 | | | |
| DB | 96 | GGAAAGCTATCTTATCTTCTTCCACGAGTCGAGTGCCTCAAAATTCATTCTGAG | 155 | | | |
| QY | 262 | GTCTGCGCCATCGGCTCACCTTTGCAAGAGAGCTTCATGTGCATAAATTTAGCAATTA | 321 | | | |
| DB | 156 | CGCTCCGCAATCATCATAGAGATTTCTTGGAATTTCTATATTAACAAGATTTGTTAAATTT | 215 | | | |
| QY | 322 | GGTTTAGGGGTGGCATTAGGTAATTCAGCAGAAATTTGGGTTAAAGAACAGAGAGAA | 381 | | | |
| DB | 216 | GATTAAAGGGGTGTGTAGCAATAATCCTTATTTGGTTCATATGTAATAAAAGGTT | 275 | | | |
| QY | 382 | TCCGCTATCCACAAATCAGAGCATTTGACAGAA | 418 | | | |
| DB | 276 | TAGGCCCTCGCAAAAAGAAAGAAATAAAAAGAA | 312 | | | |
| RESULT 10 | | | | | | |
| CA808007 | 742 bp | mRNA | linear | EST 10-APR-2003 | | |
| CA808007 | CA12L104ITF_H11 | Gabernet Sauvignon leaf - CA12L1 | Vitis vinifera | | | |
| DEFINITION | | | | | | |
| VERSION | CA808007 | | | | | |
| KEYWORDS | | | | | | |
| SOURCE | CA808007.1 | GI:26256944 | | | | |
| ORGANISM | | | | | | |
| REFERENCE | | | | | | |
| AUTHORS | | | | | | |
| TITLE | | | | | | |
| COMMENT | | | | | | |
| JOURNAL | | | | | | |
| CONTACT | | | | | | |
| CA85 | | | | | | |
| Genome Facility | | | | | | |

UC Davis Department of Plant Pathology
1 Shields Ave., Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcoc@ucdavis.edu

Seq primer: GPTATCAGTCACGGTACC.
Location/Qualifiers

FEATURES

source

1..742

/organism="Vitis vinifera"

/mol_type="mRNA"

/cultivar="Cabernet Sauvignon"

/db_xref="taxon:29760"

/clone="CA121I041IF_H11"

/sex="hermaphrodite"

/dev_stage="late season sample"

/lab_host="D5Alpha"

/clone_1ib="Cabernet Sauvignon leaf - CA121I"

/note="Organ: Leaf; Vector: pDNR; Site 1: SfiI; Site 2: SfiI; CA121I is a cDNA library of Cabernet Sauvignon leaves. The leaves were collected on September 20, 2001, in Napa Valley, California, and represent leaves in late season development. These leaves were symptomatic and verified to be infected with the bacterial pathogen, Xylella fastidiosa, based on a diagnostic assay using PCR and Xylella-specific primer pairs. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:

5'-AAGCAGTGTATCAGCAGCATGCGCTTACGGCCGG-3' and

5'-ATTCTAGAGCGCCGAGCGCGCCGACATG-3' (30'NN-3'). Library was constructed using the Clontech Creator SMART Kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 6.2%; Score 34.2; DB 14; Length 742;
Best Local Similarity 56.8%; Pred. No. 19;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 336 ACTAGTATTGATGAGATTTGGTTAAAGAACACAGCAGCATCCGCTATCCACA 395
DB 616 ATTAGGTTATTAGTTGTTATTGTTATTAAGAACTGCAATCTTAATCTTTATCCGAG 675
QY 396 AAAATCAGGCGATTTGACAGACATACCTTTAGGCGTGGCAGCCCTTT 446
DB 676 CAATACAGGCCATTTGTTAAGATGTTAGCTCAAGAGGTTCTCTCTTTT 726

RESULT 11

LOCUS BQ494475 434 bp mRNA linear EST 31-OCT-2002

DEFINITION EST03641 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.

ACCESSION BQ494475

VERSION BQ494475.1 GI:24443116

KEYWORDS EST.

SOURCE

ORGANISM

Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Omygenales; mitosporic Omygenales; Paracoccidioides.

1 (bases 1 to 434)

Goldman,G.H., dos Reis Marques,E., Duarte Ribeiro,D.C., de Souza

Bernardes,L.A., Quiapin,A.C., Vitorrelli,P.M., Savoldi,M.,

Semghini,C.P., de Oliveira,R.C., Nunes,L.R., Travassos,L.R.,

Puccia,R., Batista,W.L., Ferreira,L.E., Moreira,J.C.,

Bogossian,A.P., Tekala,F., Nobrega,M.P., Nobrega,F.G. and

Goldman,M.H.

Expressed sequence tag analysis of the human pathogen

Paracoccidioides brasiliensis yeast phase: identification of

putative homologues of Candida albicans virulence and pathogenicity

Genes

Eukaryot. Cell 2 (1), 34-48 (2003)

Contact: Gustavo Henrique Goldman

Laboratory of Molecular Biology

Universidade de Sao Paulo - USP - FCRP

Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil

Email: ggoldman@usp.br.
Location/Qualifiers

source

1..434

/organism="Paracoccidioides brasiliensis"

/mol_type="mRNA"

/db_xref="taxon:121759"

/clone_1ib="Pb0001"

ORIGIN

Query Match 6.1%; Score 33.8; DB 13; Length 434;
Best Local Similarity 54.4%; Pred. No. 20;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 429 AGGCTGCGCAGCCTTTTACTGCGGATCCACTGACCTCCAGTGGTACCGTACAGTGA 488
DB 352 AGGCTGCTCACCACGATGAGTGGCGAGGAGATTACGACAGGTAGCTCTGACACA 293
QY 489 AGAAGCAAAATAGCTGCGATGACAGATCATATCGCATACACACTTCTATACAGAA 548
DB 292 AGAAGACGCGCAGATACCGGCGTATGATAACCCCAACCAATCACCACCTATAC 233
QY 549 TGCTG 553
DB 232 CGCTG 228

RESULT 12

LOCUS BQ493352 443 bp mRNA linear EST 31-OCT-2002

DEFINITION EST02518 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.

ACCESSION BQ493352

VERSION BQ493352.1 GI:24440281

KEYWORDS EST.

SOURCE

ORGANISM

Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Omygenales; mitosporic Omygenales; Paracoccidioides.

1 (bases 1 to 443)

Goldman,G.H., dos Reis Marques,E., Duarte Ribeiro,D.C., de Souza

Bernardes,L.A., Quiapin,A.C., Vitorrelli,P.M., Savoldi,M.,

Semghini,C.P., de Oliveira,R.C., Nunes,L.R., Travassos,L.R.,

Puccia,R., Batista,W.L., Ferreira,L.E., Moreira,J.C.,

Bogossian,A.P., Tekala,F., Nobrega,M.P., Nobrega,F.G. and

Goldman,M.H.

Expressed sequence tag analysis of the human pathogen

Paracoccidioides brasiliensis yeast phase: identification of

putative homologues of Candida albicans virulence and pathogenicity

Genes

Eukaryot. Cell 2 (1), 34-48 (2003)

Contact: Gustavo Henrique Goldman

Laboratory of Molecular Biology

Universidade de Sao Paulo - USP - FCRP

Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil

Email: ggoldman@usp.br.

Location/Qualifiers

source

1..443

/organism="Paracoccidioides brasiliensis"

/mol_type="mRNA"

/db_xref="taxon:121759"

/clone_1ib="Pb0001"

ORIGIN

Query Match 6.1%; Score 33.8; DB 13; Length 443;
Best Local Similarity 54.4%; Pred. No. 20;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 429 AGGCTGCGCAGCCTTTTACTGCGGATCCACTGACCTCCAGTGGTACCGTACAGTGA 488
DB 188 AGGCTGCTCACCACGATGAGTGGCGAGGAGTTGACGAGAGGTAGCTCTGACACA 129
QY 489 AGAAGCAAAATAGCTGCGATGACAGATCATATCGCATACACACTTCTATACAGAA 548
DB 128 AGAAGACGCGCAGATACCGGCGTATGATAAACCCCAACCAATCACCACCTATAC 69

| QY | 549 | TCCTG | 553 |
|-----------------------|---|--|--------------------|
| Db | 68 | CGCTG | 64 |
| RESULT 13 | | | |
| LOCUS | B0499441/c | 485 bp | linear |
| DEFINITION | EST08666 Pb0001 Paracoccidioides brasiliensis CDNA, mRNA sequence. | | |
| ACCESSION | B0499441 | | |
| VERSION | B0499441.1 | GI:24452737 | |
| KEYWORDS | EST. | | |
| SOURCE | Paracoccidioides brasiliensis | | |
| ORGANISM | Paracoccidioides brasiliensis | | |
| REFERENCE | Paracoccidioides brasiliensis | | |
| AUTHORS | Enxayota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eukaryota; mitosporic Myxomycetes; Paracoccidioides. | | |
| | 1 (bases 1 to 485) | | |
| | Goldman G.H., dos Reis Marques E., Duarte Ribeiro D.C., de Souza | | |
| | Bernardes L.A., Quispin A.C., Vitorrelli P.M., Savoldi M., | | |
| | Semlunhni C.P., de Oliveira R.C., Nunes J.R., Travares J.R., | | |
| | Puccia R., Batista M.L., Ferreira L.E., Moreira U.C., | | |
| | Bogossian A.P., Tekala F., Nobrega M.P., Nobrega F.G. and | | |
| | Goldman M.H. | | |
| TITLE | Expressed sequence tag analysis of the human pathogen | | |
| | Paracoccidioides brasiliensis yeast phase: identification of | | |
| | putative homologues of Candida albicans virulence and pathogenicity | | |
| JOURNAL | genes | | |
| COMMENT | Eukaryot. Cell 2 (1), 34-48 (2003) | | |
| | Contact: Gustavo Henrique Goldman | | |
| | Laboratory of Molecular Biology | | |
| | Universidade de Sao Paulo - USP - FCPRP | | |
| | Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil | | |
| | Email: gsgoldman@usp.br | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..485 | | |
| | /organism="Paracoccidioides brasiliensis" | | |
| | /mol_type="mRNA" | | |
| | /db_xref="taxon:121759" | | |
| | /clone_lib="PD0001" | | |
| ORIGIN | | | |
| Query Match | 6.1% | Score 33.8; | DB 13; Length 485; |
| Best Local Similarity | 54.4%; | Pred. No. 21; | |
| Matches | 68; Conservative | 0; Mismatches | 57; Indels |
| | | | 0; Gaps |
| QY | 429 | AGGGGTGGCGAGCCCTTACTGCGGAGTCCACTGAGTCGATGGTACCGTACAGTGA | 488 |
| Db | 239 | AGGGGTGGCTACCCAGTATGTCGGGAGGAGATTACCGAGAAGTACTGCTGACAA | 180 |
| QY | 489 | AGAAACCAAAATATGCTGGCGATGACAGATCATATATGCGGATATTCACCACTTATACGAGA | 548 |
| Db | 179 | AGAAACCGCGAGATATACCGGCGTGAAGAAACCGCAACAGTACACCCACCTATAC | 120 |
| QY | 549 | TGCTG | 553 |
| Db | 119 | CGCTG | 115 |
| RESULT 14 | | | |
| LOCUS | BE776398 | 593 bp | mRNA linear |
| DEFINITION | MY-15-C-12 PlinfestansMY Phytophthora infestans cDNA, mRNA sequence. | | |
| ACCESSION | BE776398 | | |
| VERSION | BE776398.1 | GI:10230053 | |
| KEYWORDS | EST. | | |
| SOURCE | Phytophthora infestans (potato late blight agent) | | |
| ORGANISM | Phytophthora infestans | | |
| | Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; | | |
| | Phytophthora. | | |
| | 1 (bases 1 to 593) | | |
| REFERENCE | Kamoun S., Harber P.T., Sobral B.W.S., Nuss D. and Govers F. | | |
| AUTHORS | Initial assessment of gene diversity for the oomycete pathogen | | |
| TITLE | | | |

JOURNAL
MEDLINE
PUBMED
COMMENT

Phytophthora infestans based on expressed sequences
Pungel Genet. Biol. 28 (2), 94-106 (1999)
20086376
10587472
Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES
source

1. 593
/organism="Phytophthora infestans"
/mol_type="rRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DH5-alpha"
/clone_idb="PinfestansMY"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."

ORIGIN

Query Match 6.1k; Score 33.8; DB 10; Length 593;
Best Local Similarity 49.7k; Pred. No. 23;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 349 GCAGATTTGGGTTAAAGAACGACGACGATCCGCTATCCACAAATCAGGAGT 408
DB 110 GCTGAATTTAGCCACTTAGATCACCGCAATGGCAAGCGCAACGTCTCAAGTTGGGGGC 169
QY 409 ATTGCAGAAACGATACCTTTAGGGCTCGCGACCTTTTACTGCGCGATCCACTGCC 468
DB 170 ACATCCGTGGGCAATCGTCGACACTGTGGGGCTTGTACCATGTGTCGCGAGAGAGA 229
QY 469 GATGGTACCGTCGACAGTAGAAGAACCAATATGCTGCGATGACAGATCATAT 521
DB 230 GATGGCGCTTGGCGCGCTGTAGTGTCCGCAATGGGCCACACGACCATCATTT 282

RESULT 15
LOCUS BX311247/c
DEFINITION BX311247 AGENAE Rainbow trout multi-tissues subtracted library
(tcay) Oncorhynchus mykiss cDNA clone tcay0023b.h.08 3pdim, mRNA
sequence.
ACCESSION BX311247
VERSION BX311247.2 GI:40226031
SOURCE EST.
ORGANISM Oncorhynchus mykiss (rainbow trout)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Proclanchiopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 624)
Goverom, M., Guiguen, Y. and Le Gac, F.
Construction and primary characterization of normalized cDNA
libraries in rainbow trout, Oncorhynchus mykiss
unpublished (2003)
On Apr 7, 2003 this sequence version replaced gi:29591892.
Contact: Guiguen Y
INRA - SCRIBE
Campus de Beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us

at sigenasupport@joy.inra.fr to obtain the chromatogram of this sequence.

Plate: 0023 row: h column: 8
Seq primer: M13F.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| SOURCE | 1. .624 |

FEATURES

Source

1. .624

```
/organsim="Oncorhynchus mykiss"
/mol_type="RNA"
```

```
/mol_type="mRNA"
/db_xref="taxon:8
```

```
/db_xref="taxon:8022"
/cname="Tcay0023b h 0
```

```
/clone="ecay0023b.n.08"
```

differentiating gonads, gills, interrenal, /crissue_type=alprose crissue, blood, brain

kidney, liver, muscle, ovary, pituitary, testis"

```
/dev stage="from embryos to adults"
```

```
/lab_host="DH10B"
```

```
/clone_lib="AGENA
```

```
library(tcay)"
```

/note="Vector: pT7T3D-pac; Rainbow trout multi-tissues -

normalized + 1 subtraction (tday); Clone distribution:

AGENAE Resource centre. Francois PUMI,
Bureau of Development for the CEA

François.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (IREC) Domaine de Vilvert 93153

Etude du genome (LREG), Domaine de Vilvert, 78352,
Touren-Joeas cedex FRANCE

Jouy-en-Josas cedex, FRANCE

ORIGIN

Query Match 6.1%; Score 33.8; DB 13; Length 624;

Best Local Similarity 50.3%; Pred. No. 24;

| | | | | | | | | | |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|
| Matches | 83; | Conservative | 0; | Mismatches | 82; | Indels | 0; | Gaps | 0; |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|

277 TCACCTTGCAGGAGCTTCCATTGTCAATAAATGAGCAATTAGCTTAGGGGTGCA 336

Db 327 TCACATATAGATAAGTTAGATTATAATTAAATTGAATATATCATTTATCTGCTGTA 2688

337 CTAGGTAATTCAGCAGAATTGGTTAAAGAACACGACGACGAATCCCGTATCCACAA 396

Db 267 CCGGTAATCACAGTGA CTAAATAAAATTACAAGAGGAAGTAAATA GTATATACAA 208

397 AAATCAGCGAGTATTGCAGGACGATACCTTTAGGGCTGGCAGC 441

Db 207 CAAACAAGACTTTTGGTTTGGCATCCCTTACAGAGCTGTACAAC 163

Search completed: April 30, 2004, 18:35:23
Job time : 1595.49 secs

Job time : 1595.49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:32 ; Search time 65.2277 Seconds
(without alignments)
4721.887 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagattatcttctgttca.....cttcatacagagatcgtcgtc 555

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| 1 | 34.6 | 6.2 | 7218 | 1 | US-08-232-463-14 |
| 2 | 34.4 | 6.2 | 2277 | 1 | US-08-676-967-2 |
| 3 | 34.4 | 6.2 | 2277 | 1 | US-08-676-974-2 |
| 4 | 34.4 | 6.2 | 2277 | 2 | US-08-098-487-2 |
| 5 | 31.8 | 5.7 | 4715 | 4 | US-08-956-171E-203 |
| 6 | 31.6 | 5.7 | 2217 | 4 | US-09-107-532A-2564 |
| 7 | 31.6 | 5.6 | 918 | 2 | US-08-743-637B-162 |
| 8 | 31.6 | 5.6 | 918 | 3 | US-08-526-840B-162 |
| 9 | 31.6 | 5.6 | 5162 | 3 | US-09-298-367B-1 |
| 10 | 31.6 | 5.6 | 5162 | 3 | US-09-298-367B-4 |
| 11 | 31.6 | 5.6 | 5262 | 3 | US-09-298-367B-5 |
| 12 | 31.6 | 5.5 | 33248 | 4 | US-09-596-002-24 |
| 13 | 30.8 | 5.5 | 1488 | 4 | US-09-543-681A-2188 |
| 14 | 30.6 | 5.5 | 1722 | 4 | US-09-328-352-2832 |
| 15 | 30.4 | 5.5 | 805 | 4 | US-09-540-82A-16 |
| 16 | 30.4 | 5.4 | 1830121 | 4 | US-09-557-864-1 |
| 17 | 30.4 | 5.4 | 1830121 | 4 | US-09-643-990A-1 |
| 18 | 29.8 | 5.4 | 2869 | 4 | US-08-809-955-18 |
| 19 | 29.8 | 5.4 | 1664976 | 4 | US-08-916-421B-1 |
| 20 | 29.6 | 5.3 | 831 | 4 | US-08-935-263-15 |
| 21 | 29.6 | 5.3 | 831 | 4 | US-09-594-185-15 |
| 22 | 29.6 | 5.3 | 1443 | 4 | US-09-107-532A-2640 |
| 23 | 29.6 | 5.3 | 6145 | 2 | US-08-808-793-2 |
| 24 | 29.6 | 5.3 | 6145 | 2 | US-08-772-512A-2 |
| 25 | 29.6 | 5.3 | 6318 | 2 | US-08-808-793-1 |
| 26 | 29.6 | 5.3 | 6318 | 3 | US-08-772-512A-1 |
| 27 | 29.6 | 5.3 | 6318 | 3 | US-08-772-512A-1 |

| | | | | | | |
|----|------|-----|--------|---|---------------------|-------------------|
| 28 | 29.4 | 5.3 | 1239 | 4 | US-09-489-039A-6730 | Sequence 6730, Ap |
| 29 | 29.2 | 5.3 | 879 | 4 | US-09-134-000C-989 | Sequence 889, Ap |
| 30 | 29.2 | 5.3 | 1278 | 4 | US-09-491-577-45 | Sequence 45, Ap |
| 31 | 29.2 | 5.3 | 15213 | 4 | US-08-961-527-26 | Sequence 26, Ap |
| 32 | 29 | 5.2 | 3018 | 4 | US-09-328-352-751 | Sequence 751, Ap |
| 33 | 29 | 5.2 | 17310 | 4 | US-08-956-171E-23 | Sequence 23, Ap |
| 34 | 28.8 | 5.2 | 936 | 3 | US-08-910-501-3 | Sequence 3, Ap |
| 35 | 28.8 | 5.2 | 936 | 3 | US-09-398-550-3 | Sequence 3, Ap |
| 36 | 28.8 | 5.2 | 939 | 3 | US-08-910-501-1 | Sequence 1, Ap |
| 37 | 28.8 | 5.2 | 939 | 3 | US-09-398-550-1 | Sequence 1, Ap |
| 38 | 28.8 | 5.2 | 2060 | 4 | US-09-023-655-1153 | Sequence 1153, Ap |
| 39 | 28.8 | 5.2 | 41100 | 4 | US-09-755-665-46 | Sequence 46, Ap |
| 40 | 28.6 | 5.2 | 2623 | 4 | US-09-023-655-307 | Sequence 307, Ap |
| 41 | 28.6 | 5.2 | 3133 | 1 | US-08-162-809-1 | Sequence 1, Ap |
| 42 | 28.6 | 5.2 | 4860 | 4 | US-09-328-352-3221 | Sequence 3221, Ap |
| 43 | 28.6 | 5.2 | 6281 | 4 | US-09-976-594-1039 | Sequence 1039, Ap |
| 44 | 28.6 | 5.2 | 193303 | 4 | US-09-497-855A-37 | Sequence 37, Ap |
| 45 | 28.6 | 5.2 | 193303 | 4 | US-09-497-855A-44 | Sequence 44, Ap |

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZspc-F15
US-08-232-463-14
Query Match 6.2%; Score 34.6; DB 1; Length 7218;

GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Oseman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELEPHONE: (415)343-4342
TELEFAX: (415)343-4341
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-098-487-2

Query Match 6.2%; Score 34.4; DB 2; Length 2277;
Best Local Similarity 34.5%; Pred. No. 0.065;
Matches 51; Conservative 25; Mismatches 72; Indels 0; Gaps 0;

QY 11 TTCTGTCATGAGCTATATGATGCGCTGTAATCATCCGCTTATGATGATGCTGTC 70
DB 662 TTTTATTNACNSMTTCCTGRTGTTNSWYTCRTGNSMTTTCCTCCCATNCGNSWN 603
QY 71 ATMAATGGGATATGATGCTCAACCATTAAGTCACTGCGATATGATGATGAG 130
DB 602 ACNSWYTGRTGCTTATATTTTCTTTCGTCGACGCTGACGCTGACGCTGAC 543
QY 131 CCATTTTCGCGCTGATGACGCTGCT 158
DB 542 CCYTTDATYTCYTCATRTTCATNCCYT 515

RESULT 5
US-08-956-171E-203/C
Sequence 203, Application US/08956171E
GENERAL INFORMATION:
APPLICANT: Charles Kunesh
Gail H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 203:
SEQUENCE CHARACTERISTICS:
LENGTH: 4715 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 203:
US-08-956-171E-203

Query Match 5.7%; Score 31.8; DB 4; Length 4715;
Best Local Similarity 59.3%; Pred. No. 0.97;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 2 TGCAGATTTATCTGTTGATGAGCTATATGATGATGCTGTAATGATGATGCTGATG 61
DB 4275 TTATCTTCATCTTCTTCATGATGATGATGATGATGATGATGATGATGATGATG 4216
QY 62 ATGCTGCAATATGAGGTTATGCTGATCA 92
DB 4215 ATCTTCTGATCATTTAGTCTTCTTGATCA 4185

RESULT 6
US-09-107-532A-2564
Sequence 2564, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:

```

NAME: Atinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2564:
SEQUENCE CHARACTERISTICS:
LENGTH: 2217 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
TOPOLOGY: circular
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...2217
SEQUENCE DESCRIPTION: SEQ ID NO: 2564:
US-09-107-532h-2564

Query Match
Best Local Similarity 5.7%; Score 31.6; DB 4; Length 2217;
Matches 61; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2 TGCAGATATTTCTTCTTCAATGACCTATATGATGCGCTGTATGATCGCTTATGTC 61
DB 1055 TGGAAACGCTATTGTACTTATTTTAAAGATGAGTAAAGGCGAGACTTAAATC 1114
QY 62 ATGCTCGATTAATGGTATCTGCTCAACCATTAAGCTTAACTCA 111
DB 1115 AGCGTTGACGACATTTGGTTATAGTTCGAAATGACTGCTATATACA 1164

RESULT 7
US-08-743-637B-162
Sequence 162, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 918 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-743-637B-162

Query Match
Best Local Similarity 5.6%; Score 31; DB 2; Length 918;
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 310 ATTGAGCAATTAGTTAGTGGGCTGACCTGTAATTCAGCAGAAATTTGGTTAAAGAA 369
DB 624 ATTGGGCAACACATTAACCGAATCGCAAAAACGATTTGTGAAATTTGGTGGACATTA 683
QY 370 CAGACGACGAATCCGCTATCCAAATTCAGGAGTATTCAGGACATACCTTA 429
DB 684 CGCAACGAGCAATCATTGATTTGCGCTGCTACCAACTCGTGAAGATGACATTA 743
QY 430 GGGCTGGCAGCCCTTTTACTGCGCATCCACTGACTCGATGACCTGACGTAAGAA 489
DB 744 AAGCGGGCGGGTAATATGATGATGACGCAATGATTTGCGGTTCGATACCAATCG 803
QY 490 GAACCAATAGCTGCA 508
DB 804 CAACCGATTGTATGCA 822

RESULT 8
US-08-526-840B-162
Sequence 162, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 162:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 918 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-06-526-840B-162

Query Match
Best Local Similarity 47.2%; Score 31; DB 3; Length 918;
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGGTAATTGACGAGAATTGGTTAAAGAA 369
624 ATTGGGCAACACATTACCGAATGCCAAAAACGATTTTGGAATTGGTTGACATTA 683
370 CACGACGAGAAATCCCGCTATCCAAAAATCAGCACTATTGACAGAAACGATACCTTTA 429
684 CGCAACAGGCAATCCATGATTCGCGCTGCTACGCCCAATCGTGAAGTGTACGATTA 743
430 GGGCTGCGCAGCCTTTTACTGCGCATCCACTGCACTCCGATGGTACCGTACAGTAGAA 489
744 AAGCGGGCGGGTAATATGCTGTAACGCAATGATATTGGGTTGGCATACCAAAATCG 803
490 GAAACCAAAATAGCTGCA 508
804 CAACCGATTGTGATGCA 822

RESULT 9
US-09-298-367B-1/c
Sequence 1, Application US/09298367B
Patent No. 6180112

GENERAL INFORMATION:
APPLICANT: Highlander, Sarah K.
APPLICANT: Federova, Natalie D.
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE
FILE REFERENCE: BCM-03728
CURRENT APPLICATION NUMBER: US/09/298,367B
CURRENT FILING DATE: 1999-04-22
PRIORITY FILING DATE: 1997-04-15
PRIORITY FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 5162
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-298-367B-1

Query Match
Best Local Similarity 47.2%; Score 31; DB 3; Length 5162;
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGGTAATTGACGAGAATTGGTTAAAGAA 369
3719 ATTGGGCAACACATTACCGAATGCCAAAAACGATTTTGGAATTGGTTGACATTA 683
370 CACGACGAGAAATCCCGCTATCCAAAAATCAGCACTATTGACAGAAACGATACCTTTA 429
3659 CGCAACAGGCAATCCATGATTCGCGCTGCTACGCCCAATCGTGAAGTGTACGATTA 743
430 GGGCTGCGCAGCCTTTTACTGCGCATCCACTGCACTCCGATGGTACCGTACAGTAGAA 489
3599 AAGCGGGCGGGTAATATGCTGTAACGCAATGATATTGGGTTGGCATACCAAAATCG 803
490 GAAACCAAAATAGCTGCA 508
3539 CAACCGATTGTGATGCA 521

RESULT 10
US-09-298-367B-4/c
Sequence 4, Application US/09298367B
Patent No. 6180112

GENERAL INFORMATION:
APPLICANT: Highlander, Sarah K.
APPLICANT: Federova, Natalie D.
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE
FILE REFERENCE: BCM-03728
CURRENT APPLICATION NUMBER: US/09/298,367B
CURRENT FILING DATE: 1999-04-22
PRIORITY FILING DATE: 1997-04-15
PRIORITY FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 4
LENGTH: 5162
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-298-367B-4

Query Match
Best Local Similarity 47.2%; Score 31; DB 3; Length 5162;
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGGTAATTGACGAGAATTGGTTAAAGAA 369
3719 ATTGGGCAACACATTACCGAATGCCAAAAACGATTTTGGAATTGGTTGACATTA 683
370 CACGACGAGAAATCCCGCTATCCAAAAATCAGCACTATTGACAGAAACGATACCTTTA 429
3659 CGCAACAGGCAATCCATGATTCGCGCTGCTACGCCCAATCGTGAAGTGTACGATTA 743
430 GGGCTGCGCAGCCTTTTACTGCGCATCCACTGCACTCCGATGGTACCGTACAGTAGAA 489
3599 AAGCGGGCGGGTAATATGCTGTAACGCAATGATATTGGGTTGGCATACCAAAATCG 803
490 GAAACCAAAATAGCTGCA 508
3539 CAACCGATTGTGATGCA 521

RESULT 11
US-09-298-367B-5/c
Sequence 5, Application US/09298367B
Patent No. 6180112

GENERAL INFORMATION:
APPLICANT: Highlander, Sarah K.
APPLICANT: Federova, Natalie D.
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE
FILE REFERENCE: BCM-03728
CURRENT APPLICATION NUMBER: US/09/298,367B
CURRENT FILING DATE: 1999-04-22
PRIORITY FILING DATE: 1997-04-15
PRIORITY FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 5262
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-298-367B-5

Query Match
Best Local Similarity 47.2%; Score 31; DB 3; Length 5262;
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGGTAATTGACGAGAATTGGTTAAAGAA 369

Db 4158 ATTGGCACAACATTAACCGAATCGCAAAAAAGATTTTGTGGAAATGGTTGGACAATA 4099
QY 370 CACGACGACGAATCCCGCTATCCACAAAATCAGGCGATTTGCAGGAACGATACCTTTA 429
Db 4098 CGGACAGGCAATCCATGATTCGCGCTCTAGCGCAACATGCTGGAAGTGTAGATTA 4039
QY 430 GGGCTCGCAGCCTTTTACTGGCGCATCGACCTGGAATCCGATGTAACCGTACAGTA 489
Db 4038 AAGCGGGCGGGTAATATGTGTAGCAATGATATGGGTGGTTGGATACCAATCG 3979
QY 490 GAAACCAAAATAGCTGGCA 508
Db 3978 CAAACCGATTGTGATGCA 3960

RESULT 12
US-09-298-367B-2/c
; Sequence 2, Application US/09298367B
; Patent No. 6180112
; GENERAL INFORMATION:
; APPLICANT: Highlander, Sarah K.
; APPLICANT: Federova, Natalie D.
; TITLE OF INVENTION: PASTURELHA HAEMOLYTICA VACCINE
; FILE REFERENCE: BCM-03728
; CURRENT APPLICATION NUMBER: US/09/298,367B
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 08/834,455
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6850
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-298-367B-2

Query Match 5.6%; Score 31; DB 3; Length 6850;
Best Local Similarity 47.2%; Pred. No. 2.4;
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 310 ATTGACCAATTAGTTTAAAGGCTGGACATGATTAATCGACGAATTTGGTTAAAGAA 369
Db 5407 ATTGGCAACATTAACCGAATCGCAAAAAGATTTTGGAAATGGTTGGACAATTA 5348
QY 370 CACGACGACGAATCCCGCTATCCACAAAATCAGGCGATTTGCAGGAACGATACCTTTA 429
Db 5347 CGGACAGGCAATCCATGATTCGCGCTCTAGCGCAACATGCTGGAAGTGTAGATTA 5288
QY 430 GGGCTCGCAGCCTTTTACTGGCGCATCGACCTGGAATCCGATGTAACCGTACAGTA 489
Db 5287 AAGCGGGCGGGTAATATGTGTAGCAATGATATGGGTGGTTGGATACCAATCG 5228
QY 490 GAAACCAAAATAGCTGGCA 508
Db 5227 CAAACCGATTGTGATGCA 5209

RESULT 13
US-09-596-002-24/c
; Sequence 24, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18

NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 33248
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 24
; PUBLICATION INFORMATION:
US-09-596-002-24

Query Match 5.5%; Score 30.8; DB 4; Length 33248;
Best Local Similarity 45.2%; Pred. No. 8.4;
Matches 113; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 138 TCGCCGCTTGACCGATGCTCACTGCTGCTGCTTAATGCTTGTGACACAGTTT 197
Db 11987 TCCATGACCGCCGCCAAGCAGCATTTACGCCAGCTTACCGCTGCTGCTGCTT 11928
QY 198 GGGCGATTGTGTATCAACGTTATGAAATCGCGGACCGCTGCTGGAACCTCTC 257
Db 11927 TGGCAAGGGGCAATTAATCATCTGTGATTTTAAAGTACGCTCGATTAACAAAC 11868
QY 258 CCATGTGCGCATCGGCTCACTGCTGCAAGAGCTTCCATTTGCAATTAATGAGCA 317
Db 11867 GCTTTATGTCATGATGCTTATCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 11808
QY 318 ATTAGTTTGGGCTGACATGATTAATGACGAAATTTGGTTAAAGAACAGACA 377
Db 11807 AATGTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11748
QY 378 CGAATCCCGC 387
Db 11747 GAATTTATGC 11738

RESULT 14
US-09-543-681A-2188
; Sequence 2188, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2188
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2188

Query Match 5.5%; Score 30.6; DB 4; Length 1488;
Best Local Similarity 49.7%; Pred. No. 1.2;
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 209 TGATCAACGTTATCTPAAATCGCGGACCGCTGCTGTAACCTCTCCATGCTGCTG 268
Db 1043 TGACCAAAAAGGATGATTTGCTGAAAACGAATGAATGTTCTTTTCAAGTGG 1102
QY 269 CCATGCGCTCACTTTGCAAGAGCTTCCATTTGTCAATAAATTTAGCAATTAGTTAG 328
Db 1103 AAGCGGCTCATGATTAACCAAGATCATATGCTATCTTGTGTCAATTTGGTGA 1162
QY 329 GGGTGCATAGGTATTTAGCAAGAAATTTGGTTAA 365
Db 1163 CGGTGCAATTAATGAAGCGGTACAAATGGTTTGA 1199

RESULT 15
 US-09-328-352-2832/c
 ; Sequence 2832 Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 2832
 ; LENGTH: 1722
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-2832

Query Match 5.5%; Score 30.4; DB 4; Length 1722;
 Best Local Similarity 47.4%; Pred. No. 1.6;
 Matches 91; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 322 | GGTTAGGGGGTGGCACTAGGTATTGACGAGAAATTTGGTTAAAGAAACACGACGACGAA | 381 |
| | | | |
| DB | 280 | GGTTAATGTAGTCCATGCTTGCATATAAAATAGCAAGATAGCCACTGCAGAAAAA | 221 |
| | | | |
| QY | 382 | TCCCGCTATCCACAAAATCAGGCGATGATGAGGAAGATACCTTTAGGGCTGGCGAC | 441 |
| | | | |
| DB | 220 | TACAGCTCACAACAAATGCTAGAGAAATGTTTCCAGTTCTTTGTGTGTGGAGATA | 161 |
| | | | |
| QY | 442 | CTTTACTGCGCGATCCACTGGAATCGATGATGATCCGTACAGTAGAAGAAACCAATA | 501 |
| | | | |
| DB | 160 | CACCTAATTAAGGGTATACAGGATTTTCAATCCAGCAAAATATCATCACCCACATCG | 101 |
| | | | |
| QY | 502 | GCTGGCATGACA | 513 |
| | | | |
| DB | 100 | GCTGATTGCCA | 89 |

Search completed: April 30, 2004, 18:43:19
 Job time : 76.2277 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 15:29:25 ; Search time 246.227 Seconds
(without alignments)
10195.950 Million cell updates/sec

Title:

US-10-603-260-4

Perfect score:

555

Sequence:

1 atgcagattatctgtctca.....cttcacagagatctgtt 555

Scoring table:

IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched:

2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters:

5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----------------------|--------------------|
| 1 | 555 | 100.0 | 555 | US-10-603-260-4 | Sequence 4, Appl1 |
| 2 | 555 | 100.0 | 2578 | US-10-603-260-1 | Sequence 1, Appl1 |
| 3 | 36.8 | 6.6 | 26997 | US-10-221-714A-473 | Sequence 473, Appl |
| 4 | 33.2 | 6.0 | 32185 | US-10-087-192-1135 | Sequence 1135, Ap |
| 5 | 33 | 5.9 | 435 | US-09-783-590-1460 | Sequence 1460, Ap |
| 6 | 33 | 5.9 | 96499 | US-10-087-192-2011 | Sequence 2011, Ap |
| 7 | 32.8 | 5.9 | 948 | US-10-282-122A-15368 | Sequence 15368, A |
| 8 | 32.4 | 5.8 | 1691139 | US-10-067-514-1 | Sequence 1, Appl1 |
| 9 | 32.4 | 5.8 | 1691139 | US-10-419-723-1 | Sequence 1, Appl1 |
| 10 | 32.2 | 5.8 | 653122 | US-10-087-192-226 | Sequence 226, App |
| 11 | 32 | 5.8 | 549 | US-10-282-122A-18072 | Sequence 18072, A |
| 12 | 31.8 | 5.7 | 969 | US-09-938-842A-1790 | Sequence 1790, App |
| 13 | 31.8 | 5.7 | 969 | US-09-934-455-271 | Sequence 271, App |
| 14 | 31.8 | 5.7 | 969 | US-09-938-842A-1790 | Sequence 1790, App |

| | | | | | | |
|----|------|-----|--------|----|----------------------|-----------------------------------|
| 15 | 31.8 | 5.7 | 969 | 13 | US-10-225-066A-377 | Sequence 377, App |
| 16 | 31.8 | 5.7 | 969 | 16 | US-10-374-780A-2733 | Sequence 2733, Ap |
| 17 | 31.8 | 5.7 | 4715 | 8 | US-08-781-986A-203 | Sequence 203, App |
| 18 | 31.8 | 5.7 | 4715 | 13 | US-10-329-624-203 | Sequence 203, App |
| 19 | 31.6 | 5.7 | 521 | 13 | US-10-424-599-50022 | Sequence 50022, A |
| 20 | 31.6 | 5.7 | 811 | 15 | US-10-184-644-414 | Sequence 414, App |
| 21 | 31.6 | 5.7 | 811 | 15 | US-10-184-644-414 | Sequence 414, App |
| 22 | 31.6 | 5.7 | 17286 | 9 | US-09-764-877-3234 | Sequence 3234, Ap |
| 23 | 31.6 | 5.7 | 17286 | 16 | US-10-243-515-3234 | Sequence 3234, Ap |
| 24 | 31.2 | 5.6 | 612 | 13 | US-10-027-632-125951 | Sequence 125951, Sequence 125952, |
| 25 | 31.2 | 5.6 | 612 | 13 | US-10-027-632-125952 | Sequence 125951, |
| 26 | 31.2 | 5.6 | 612 | 16 | US-10-027-632-125951 | Sequence 125952, |
| 27 | 31.2 | 5.6 | 612 | 16 | US-10-027-632-125952 | Sequence 251, App |
| 28 | 31.2 | 5.6 | 1300 | 10 | US-09-934-455-251 | Sequence 251, App |
| 29 | 31.2 | 5.6 | 1300 | 13 | US-10-225-066A-841 | Sequence 841, App |
| 30 | 31.2 | 5.6 | 1300 | 16 | US-10-374-780A-309 | Sequence 309, App |
| 31 | 31.2 | 5.6 | 1308 | 13 | US-10-282-122A-14107 | Sequence 14107, A |
| 32 | 31.2 | 5.6 | 11485 | 15 | US-10-205-219-144 | Sequence 1475, Ap |
| 33 | 31.2 | 5.6 | 15275 | 9 | US-09-764-865-1475 | Sequence 1475, Ap |
| 34 | 31.2 | 5.6 | 15275 | 15 | US-10-091-504-1475 | Sequence 1475, Ap |
| 35 | 31.2 | 5.6 | 15275 | 16 | US-10-227-577-1475 | Sequence 1475, Ap |
| 36 | 31.2 | 5.6 | 856 | 13 | US-10-027-632-5858 | Sequence 5858, Ap |
| 37 | 31 | 5.6 | 856 | 16 | US-10-027-632-5858 | Sequence 5858, Ap |
| 38 | 31 | 5.6 | 918 | 9 | US-09-452-599-162 | Sequence 162, App |
| 39 | 31 | 5.6 | 918 | 15 | US-10-121-120-162 | Sequence 162, App |
| 40 | 31 | 5.6 | 105184 | 10 | US-09-847-513A-1 | Sequence 1, Appl1 |
| 41 | 30.8 | 5.5 | 33248 | 13 | US-10-672-787-24 | Sequence 24, Appl1 |
| 42 | 30.6 | 5.5 | 476 | 9 | US-09-864-761-5512 | Sequence 5512, App |
| 43 | 30.6 | 5.5 | 659 | 13 | US-10-027-632-152659 | Sequence 152659, Sequence 152659, |
| 44 | 30.6 | 5.5 | 659 | 16 | US-10-027-632-152659 | Sequence 32888, A |
| 45 | 30.6 | 5.5 | 1437 | 13 | US-10-282-122A-32888 | |

ALIGNMENTS

RESULT 1
US-10-603-260-4
Sequence 4, Application US/10603260
Publication No. US20040009570A1
GENERAL INFORMATION:
APPLICANT: CJ Corporation
TITLE OF INVENTION: An alkaline lipase from Vibrio metchnikovi RH530 and a
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/603,260
CURRENT FILING DATE: 2003-06-24
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Kopatentim 1.71
SEQ ID NO 4
LENGTH: 555
TYPE: DNA
ORGANISM: Vibrio metchnikovi RH530
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(555)
OTHER INFORMATION: val12 gene
US-10-603-260-4

Query Match 100.0%; Score 555; DB 16; Length 555;
Best Local Similarity 100.0%; Pred. No. 1.6e-179;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGATTATCTGTCTCATGACTCTATGATGATGCTTGTGATGATGATGCTTAT 60
DB 1 ATGCAGATTATCTGTCTCATGACTCTATGATGATGCTTGTGATGATGATGCTTAT 60
QY 61 CATGCGTCGATTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 CATGCGTCGATTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 GATGATGAGGATTTTGGCGGCTTGAACGATGCTCATGCTGCTGCTGCTATGCT 180
DB 121 GATGATGAGGATTTTGGCGGCTTGAACGATGCTCATGCTGCTGCTGCTATGCT 180

| | | | |
|----|-----|---|-----|
| Db | 12 | GATGATGAGGCCATTTTTGCCGCCCTTGACGATCGCTCATCTCANTGCTCGCCTTAAGCT | 180 |
| Qy | 181 | TTAGTGGGACACAGCTTTGGGCGGAGTTGGTGAATCAAAAGTTATCTAGAAATGGCGCGACCG | 240 |
| Db | 181 | TTAGTGGGACACAGTTTGGCGGAGTTGGTGAATCAAAAGTTATCTAGAAATGGCGCGACCG | 240 |
| Qy | 241 | TCCTGTGAACCCCTCTCCCATGTCGTGCGCATCGGCTCACTTTGCAAGAGCTTCATT | 300 |
| Db | 241 | TCCTGTGAACCCCTCTCCCATGTCGTGCGCATCGGCTCACTTTGCAAGAGCTTCATT | 300 |
| Qy | 301 | GTCATATAAATTGAGCAATTAGGTTTAAAGGTTGCGACTAGGTAATTGACAGAAATTTGGG | 360 |
| Db | 301 | GTCATATAAATTGAGCAATTAGGTTTAAAGGTTGCGACTAGGTAATTGACAGAAATTTGGG | 360 |
| Qy | 361 | TTTAAAGAACAGACGACGATTCGCGCTATCCACAAAATAGGACGATTTGACAGAAAG | 420 |
| Db | 361 | TTTAAAGAACAGACGACGATTCGCGCTATCCACAAAATAGGACGATTTGACAGAAAG | 420 |
| Qy | 421 | ATACTTTAAGGCGTCGCGACGCTTTTATCTGCGCGATCCACTGGACTCCGATGTGTAACGTC | 480 |
| Db | 421 | ATACTTTAAGGCGTCGCGACGCTTTTATCTGCGCGATCCACTGGACTCCGATGTGTAACGTC | 480 |
| Qy | 481 | ACAGTAGAAGAAAACCAAATAGCTGGCATGACAGATCATATCGGAGATTCACACATTTCA | 540 |
| Db | 481 | ACAGTAGAAGAAAACCAAATAGCTGGCATGACAGATCATATCGGAGATTCACACATTTCA | 540 |
| Qy | 541 | TACGAGATGCTGTT | 555 |
| Db | 541 | TACGAGATGCTGTT | 555 |

```

RESULT 2
US-10-603-260-1
; Sequence 1, Application US/10603260
; Publication No. US20040009570A1
; GENERAL INFORMATION:
; APPLICANT: CT Corporation
; TITLE OF INVENTION: An alkaline lipase from Vibrio metchnikovii RH530 and a
; FILE OF INVENTION: nucleotide sequence encoding the same
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/603,260
; CURRENT FILING DATE: 2003-06-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Kopacenta 1.71
; SEQ ID NO 1
; LENGTH: 2578
; TYPE: DNA
; ORGANISM: Vibrio metchnikovii RH530
US-10-603-260-1

```

| | | | | | | | | | | |
|----|-----------------------|---|--------------|-------|------------|----|--------|--------|------|---|
| | Query Match | Similarity | 100.0% | Score | 555 | DB | 16 | Length | 2578 | |
| | Best Local Similarity | 100.0% | Pred. | No. | 4e-179 | | | | | |
| | Matches | 555 | Conservative | 0 | Mismatches | 0 | Indels | 0 | Gaps | 0 |
| OY | 1 | ATGCAGATTATCTTGTTGTCATGGACTCATATGATCAGTGGCTGTGGTAATGCATCCGTTAGT | 60 | | | | | | | |
| Dd | 1561 | ATGCAGATTATCTTGTTGTCATGGACTCATATGATCAGTGGCTGTGGTAATGCATCCGTTAGT | 1620 | | | | | | | |
| OY | 61 | CATCGTCGCAATAATGGGTATCGTACCTAAACCATTAGTCAACACTCACTCGCTATC | 120 | | | | | | | |
| Dd | 1621 | CATCGTCGCAATAATGGGTATCGTACCTAAACCATTAGTCAACACTCACTCGCTATC | 1680 | | | | | | | |
| OY | 121 | GATGATGAGGCCATTTTTGCGCGCTTGAACGATGCTGCACTCAATGCTCGGCTTAATGCT | 180 | | | | | | | |
| Dd | 1681 | GATGATGAGGCCATTTTTGCGCGCTTGAACGATGCTGCACTCAATGCTCGGCTTAATGCT | 1740 | | | | | | | |
| OY | 181 | TTAGTCGACACAGTTTGCGCGGATGTGGATCAACAACGTATCTGAATTCGCGCACCG | 240 | | | | | | | |
| Dd | 1741 | TTAGTCGACACAGTTTGCGCGGATGTGGATCAACAACGTATCTGAATTCGCGCACCG | 1800 | | | | | | | |
| OY | 241 | TCCGTGTAAACCCCTCTCCCATGCGTCATCGTCGCAATCGGCTCAACCTTGTSCAAGAAGCTTCATT | 300 | | | | | | | |
| Dd | 1801 | TCCGTGTAAACCCCTCTCCCATGCGTCATCGTCGCAATCGGCTCAACCTTGTSCAAGAAGCTTCATT | 1860 | | | | | | | |

| | | | |
|----|------|--|------|
| QY | 301 | GTCAATAAAATTAGCAATTAAGTTTAGGGGTGGCACTAGGTATTTGACGAAATTTGGG | 360 |
| Db | 1861 | GTCAATTAATTTAGCAATTAGTTTAGGGGTGGCACTAGGTATTTGACGAAATTTGGG | 1920 |
| QY | 361 | TTAAAGAACAAGCAAGCAATCCGCTATTCACAAAATACGAGGAGTATTGCAAGAAC | 420 |
| Db | 1921 | TTAAAGAACAAGCAAGCAATCCGCTATTCACAAAATACGAGGAGTATTGCAAGAAC | 1980 |
| QY | 421 | ATTACCTTTAGGGGTGGCGAGGCTTTTACTGGCGCATCCATGGACTCGAATGTAACGTC | 480 |
| Db | 1981 | ATTACCTTTAGGGGTGGCGAGGCTTTTACTGGCGCATCCATGGACTCGAATGTAACGTC | 2040 |
| QY | 481 | ACAGTAGAAGAAACCAAAATATGCTGGCATGACAGATCATATCGGATATCCACCACTTCA | 540 |
| Db | 2041 | ACAGTAGAAGAAACCAAAATATGCTGGCATGACAGATCATATCGGATATCCACCACTTCA | 2100 |
| QY | 541 | TACAGAGATGCTGTT | 555 |
| Db | 2101 | TACAGAGATGCTGTT | 2115 |

```

RESULT 3
US-10-221-714A-473/c
; Sequence 473. Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221.714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO: 473
; LENGTH: 26897
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-473

```

| | Query Match | Best Local Similarity | 4.6%; | Score 36.8; | DB 13; | Length 26997; |
|----|--------------|---|--------|----------------|--------|-------------------|
| | | | 4.71%; | Pred. No. 1.4; | | |
| | Matches 113; | Conservative | 0; | Mismatches | 127; | Indels 0; Gaps 0; |
| QY | 277 | TCACCTTTGCAGAGAGCTTCATGTCATATAAATTGAGCAATTTAGGTTAGGGGTGGCA | 356 | | | |
| DB | 13416 | TCTCCATTCGAAACCTAACACATCCATTAATACAAATTAAACACATTAATATTC | 1335 | | | |
| QY | 337 | CTAGGTAAATCAGAGAAATTGGGTTAAAGAAACAGAGACGAATCCGCTATCCACAA | 396 | | | |
| DB | 13386 | CTAACACATTTTATTAACCTTGACACATTAATCCACTGGAACCAATATCCTTTAAACA | 1329 | | | |
| QY | 397 | AAATCAGGACGATATTGCGAGAACGATACCTTTAGGGGTGGCAGCCCTTTATCTGGCGAT | 456 | | | |
| DB | 13286 | ATAAAAATTTAAAAAATCAAAAAACGCTACCTTATATCTACAAAAATATATCCCGACCTCTAC | 1323 | | | |
| QY | 457 | CCACTGGACCTCCGATGGTACCGTCACAGTGAAGAAACCAAAATATAGCTGGCATGACGAT | 516 | | | |

;; CURRENT APPLICATION NUMBER: US/10/087,192
;; PRIOR FILING DATE: 2002-03-01
;; PRIOR APPLICATION NUMBER: US 09/747,377
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: US 09/798,586
;; PRIOR FILING DATE: 2001-03-02
;; NUMBER OF SEQ ID NOS: 2059
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2011
;; LENGTH: 96499
;; TYPE: DNA
;; ORGANISM: Mus musculus
;; FEATURE:
;; NAME/KEY: misc.feature
;; LOCATION: (1)...(96499)
;; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-2011

Query Match 5.9%; Score 33; DB 13; Length 96499;
Best Local Similarity 52.6%; Pred. No. 62;
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 26 TCATATGATGGCTTGAATGCAATCCGCTTAGTCATCGCTGATTAATGGTTATC 85
DB 9513 TCACACGCTTAGCAGCAATCTCTGTCTAGTTGAACCATACACAGTTTGGTTGCT 9454
QY 86 GACTCAACCACTTAGCTCAACTCACTCGCTATCGATGATGAGGCCATTTTCGCCGCC 145
DB 9453 GGACACAGAGCAATTTCTCAACTAATAGGCAATGACATGCGCTTTTTCACCT 9344
QY 146 TTGACCGATCGCTCACT 162
DB 9393 TCTACTACCCAGAGT 9377

RESULT 7
US-10-282-122A-15368
;; Sequence 15368; Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangsu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636

;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 15368
;; LENGTH: 948
;; TYPE: DNA
;; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15368

Query Match 5.9%; Score 32.8; DB 13; Length 948;
Best Local Similarity 46.8%; Pred. No. 4.7;
Matches 103; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 336 ACTAGGTAATTCAGCAAGATTGGTTAAAGACGACGAGCAATCCGCTATCCACA 395
DB 561 AATTAAGTATTAGCAAACTCACCATAGAGATGGGAATGTACTATTAAGACCAGA 620
QY 396 AAAATCAGCAGTATTGCAAGAACGATACCTTTAGGGCTGCGAGCCTTTTACTGCGCA 455
DB 621 AGAAGCAATTAGTTGGCAGCAAGAACTATTATAGACATTTTAATTATTTAGACTCT 680
QY 456 TCCACTGACTCCGATGATACCTCAAGTGAAGAAACCAAAATGCTGCATGACGA 515
DB 681 TACAGATCAAGCAGATGATGTGAGATTAATGTGAAGAAAGATTAAGAGAAA 740
QY 516 TCATATGCGATATCCACCACTTCATACGAGATGCTGT 555
DB 741 GGTTCCTGAGATGATCATGCAAGAACTGATCTTCAATT 780

RESULT 8
US-10-067-514-1/c
;; Sequence 1, Application US/10067514
;; Publication No. US20030054531A1
;; GENERAL INFORMATION:
;; APPLICANT: Gietarsdotter, Solveig
;; APPLICANT: Jonsdotter, Sif
;; APPLICANT: Reymondotter, Sigridur Th.
;; TITLE OF INVENTION: HUMAN STROKE GENE
;; FILE REFERENCE: 2345.2010-003
;; CURRENT APPLICATION NUMBER: US/10/067,514
;; CURRENT FILING DATE: 2002-02-04
;; PRIOR APPLICATION NUMBER: US 09/811/352
;; PRIOR FILING DATE: 2001-03-19
;; NUMBER OF SEQ ID NOS: 84
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1
;; LENGTH: 1691139
;; TYPE: DNA
;; ORGANISM: Human
US-10-067-514-1

Query Match 5.8%; Score 32.4; DB 15; Length 1691139;
Best Local Similarity 50.0%; Pred. No. 4.9e+02;
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 274 GGTCACTTTGCAAGAGCTTCCATTTCAAAATTAATGAGCAATAGGCTG 333
DB 1602662 GGCTCATCTTACATATTTCTTCTGAGTCCCTAGAAATCAGTATTTTTCAAAAG 1602603
QY 334 GCACTAGTAATTCAGCAAGATTGGTTAAAGAACACGACGAAATCCGCTATCCA 393
DB 1602602 TCTTGTCTCTTTTACAGAGATGATTAAGAAACCAAAATAGGCTGTAGTGCT 1602543
QY 394 CAAAATCAGCAGTATTGCAAGAACATTCCTTTAGGGCTG 435
DB 1602542 CATTTGCTACTGGGGTATGTAAGCTTTCAGCCCTCTGGCTG 1602501

RESULT 9

```
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
; APPLICANT: Gretaodotitir, Solveig
; APPLICANT: Jonadotitir, Sif
; APPLICANT: Reynadotitir, Sigridur Th.
; APPLICANT: Thorleifsson, Gudmar
; APPLICANT: Gulcher, Jeffrey
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; FILE REFERENCE: 2345.2010-005
; CURRENT APPLICATION NUMBER: US/10/419,723
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/255,120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-419-723-1

Query Match      5.8%; Score 32.4; DB 16; Length 1691139;
Best Local Similarity 50.0%; Pred. No. 4.9e+02;
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 274 GGCCTACCTTGCAAGAGAGCTTCATTCATTAATAGCATTTAGGGGCTG 333
Db 1602662 GGCCTACCTTGCAAGAGAGCTTCATTCATTAATAGCATTTAGGGGCTG 1602603

Qy 334 GCACCTAGTAATTCAGCAAGATTGGGTTAAAGAACGACGACGAATCCCGCTATCCA 393
Db 1602602 TCTTGTCCTTTTACGAGAAATGGTATTAAGAAACCAATGAGGGTGAAGTGTGCT 1602543

Qy 394 CAAATAGGAGGAGTATTCAGCAAGCATTACCTTTAGGGGCTG 435
Db 1602542 CATTCCTACCTGGGATATGATGCTTTTACACCTCTCGGCTG 1602501

RESULT 10
US-10-087-192-226/c
; Sequence 226, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226
; LENGTH: 653122
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(653122)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-226
```

```
Query Match      5.8%; Score 32.2; DB 13; Length 653122;
Best Local Similarity 61.2%; Pred. No. 3.5e+02;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 362 TAAAGACGACGACGACGATCCGCTATCCAAATAATCAGGAGTATTGACGAGACGA 421
Db 142345 TTAAGTACCAAGTACGAAATCTTCGATAGGCTTAACATCAGGACGAGTGAAGTGA 142286

Qy 422 TACCTTAGGCTGGCGACCTTTT 446
Db 142285 TAGCTTAGGATACCACTATTTT 142261

RESULT 11
US-10-282-122A-18072
; Sequence 18072, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18072
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
US-10-282-122A-18072

Query Match      5.8%; Score 32; DB 13; Length 549;
Best Local Similarity 65.3%; Pred. No. 6.3; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 332 TGGCACTAGTAATTCAGCAAGATTGGGTTAAAGAACGACGACGATCCCGCTATC 391
Db 428 TGAATCTGGAAATCTCGAAATTTTGTCGTAATTAACGATGAAGAGTCAGCTTTC 487

Qy 392 CACAAATATCAG 403
```

DB 488 TAAACAATCCG 499

RESULT 12

US-09-938-842A-1790
Sequence 1790, Application US/09938842A

Patent No. US20020160378A1
GENERAL INFORMATION:

APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SAME, AND METHODS OF USE

CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1790

LENGTH: 969
TYPE: DNA

ORGANISM: Arabidopsis thaliana
US-09-938-842A-1790

Query Match
Best Local Similarity 54.8%; Score 31.8; DB 9; Length 969;
Pred. No. 10; Mismatches 52; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

DB 89 CTCMAACCTAGTACACTCACTCGCTATCATGATGAGGCCATTTTGGCGCCTTG 148

DB 392 CTCMAACCAACCAACCAACCTCCCTGTTCTCCCTGGAAGACGGTGTCCAGCACAA 451

QY 149 ACCGATCGTCACTCACTGCTGCTTATGCTTTGTGCGACACAGTTGGCGG 203

DB 452 AACGCCCTGCTCCACTGTCAGATGAAGATGAAGTCAAGACATGTGTGCGG 506

RESULT 13

US-09-934-455-271

Sequence 271, Application US/09934455
Publication No. US20030121070A1

GENERAL INFORMATION:

APPLICANT: Adam, Luc
APPLICANT: Creelman, Robert

APPLICANT: Dubell, Arnold
APPLICANT: Heard, Jacqueline

APPLICANT: Jiang, Cai-Zhong
APPLICANT: Keddle, James

APPLICANT: Pilgrim, Marsha
APPLICANT: Ratcliffe, Oliver

APPLICANT: Reuber, Lynne
APPLICANT: Riechmann, Jose Luis

APPLICANT: Yu, Guo-Liang
APPLICANT: Yin, Omita

TITLE OF INVENTION: Genes for Modifying Plant Traits IV
FILE REFERENCE: MBI-0025

CURRENT APPLICATION NUMBER: US/09/934,455
CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/227439
PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: MBI-0022
PRIOR FILING DATE: 2001-11-16

PRIOR APPLICATION NUMBER: MBI-0023
PRIOR FILING DATE: 2001-04-17

NUMBER OF SEQ ID NOS: 516
SOFTWARE: PatentIn version 3.1
SEQ ID NO 271

LENGTH: 969
TYPE: DNA

ORGANISM: Arabidopsis thaliana
FEATURE:

NAME/KEY: CDS
LOCATION: (1)..(969)

OTHER INFORMATION: G1835
US-09-934-455-271

Query Match
Best Local Similarity 54.8%; Score 31.8; DB 10; Length 969;
Pred. No. 10; Mismatches 52; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

DB 89 CTCMAACCTAGTACACTCACTCGCTATCATGATGAGGCCATTTTGGCGCCTTG 148

DB 392 CTCMAACCAACCAACCAACCTCCCTGTTCTCCCTGGAAGACGGTGTCCAGCACAA 451

QY 149 ACCGATCGTCACTCACTGCTGCTTATGCTTTGTGCGACACAGTTGGCGG 203

DB 452 AACGCCCTGCTCCACTGTCAGATGAAGATGAAGTCAAGACATGTGTGCGG 506

RESULT 14

US-09-938-842A-1790

Sequence 1790, Application US/09938842A
Publication No. US20040009476A3

GENERAL INFORMATION:

APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SAME, AND METHODS OF USE

CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1790

LENGTH: 969
TYPE: DNA

ORGANISM: Arabidopsis thaliana
US-09-938-842A-1790

Query Match
Best Local Similarity 54.8%; Score 31.8; DB 11; Length 969;
Pred. No. 10; Mismatches 52; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 89 CTCMAACCTAGTACACTCACTCGCTATCATGATGAGGCCATTTTGGCGCCTTG 148

DB 392 CTCMAACCAACCAACCAACCTCCCTGTTCTCCCTGGAAGACGGTGTCCAGCACAA 451

QY 149 ACCGATCGTCACTCACTGCTGCTTATGCTTTGTGCGACACAGTTGGCGG 203

DB 452 AACGCCCTGCTCCACTGTCAGATGAAGATGAAGTCAAGACATGTGTGCGG 506

RESULT 15

US-10-225-066A-377

Sequence 377, Application US/10225066A
Publication No. US20030226173A1

GENERAL INFORMATION:

APPLICANT: Merdel Biotechnology, Inc.
APPLICANT: Ratcliffe, Oliver

APPLICANT: ADAM, Luc J
APPLICANT: DUBELL, Arnold T

APPLICANT: HEARD, Jacqueline E

APPLICANT: PILGRIM, Marsha L
APPLICANT: JIANG, Cai-Zhong
APPLICANT: REUBER, T. Lynne
APPLICANT: CREELMAN, Robert A
APPLICANT: PINEDA, Omaira
APPLICANT: YU, Guo-Liang
APPLICANT: BROUN, Pierre E
TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
FILE REFERENCE: MB10036-2 US
CURRENT APPLICATION NUMBER: US/10/225,066A
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: 09/837,444
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/310,847
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/336,049
PRIOR FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 60/338,692
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: 10/171,468
PRIOR FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 1122
SOFTWARE: PatentIn version 3.1
SEQ ID NO 377
LENGTH: 969
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-10-225-066A-377

Query Match 5.7%; Score 31.8; DB 13; Length 969;
Best Local Similarity 54.8%; Pred. No. 10; Mismatches 52; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 89 CTCAACCATTAAGCTACACTCACTCGCTATCGATGATGAGGCCATTTCGCCGCTTG 148
DB 392 CTCAACCATTAAGCTACACTCACTCGCTATCGATGATGAGGCCATTTCGCCGCTTG 148
QY 149 ACCGATGCTCTCACTCACTCGCTCGCTTAATGCTTTAGTGGACACAGTTGGGCGG 203
DB 452 AACGCCCTCGTCCACCTGTCAGGATTAAGATAGATCAAGACAATGTGTGCGG 506

Search completed: April 30, 2004, 23:19:59
Job time : 261.561 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:16:06 ; Search time 57 Seconds
(without alignments)
917.040 Million cell updates/sec

Title: US-10-603-260-5
Perfect score: 936
Sequence: 1 MQLIVHGLVHGLVHPLS.....KIAMGDHIAISTSYENAV 185

Scoring table: BLASTNM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 323.5 | 34.6 | 201 | 2 | AAW23078 LA11.1 es |
| 2 | 98 | 10.5 | 323 | 3 | AAW20858 Arabidops |
| 3 | 98 | 10.5 | 339 | 3 | AAW20857 Arabidops |
| 4 | 98 | 10.5 | 340 | 3 | AAW20856 Arabidops |
| 5 | 97 | 10.4 | 277 | 1 | AAW20065 Sequence |
| 6 | 97 | 10.4 | 277 | 1 | AAW48085 P. firsagi |
| 7 | 96.5 | 10.3 | 453 | 6 | ADA33782 Actinocoba |
| 8 | 96 | 10.3 | 316 | 7 | ADG94967 E. faeciu |
| 9 | 96 | 10.3 | 319 | 1 | AAW91951 Polypepti |
| 10 | 96 | 10.3 | 319 | 2 | AAW88013 Mature Ps |
| 11 | 96 | 10.3 | 319 | 2 | AAW88015 Mature Ps |
| 12 | 96 | 10.3 | 319 | 2 | AAW88018 Mature Ps |
| 13 | 96 | 10.3 | 319 | 2 | AAW88009 Mature Ps |
| 14 | 96 | 10.3 | 319 | 2 | AAW88014 Mature Ps |
| 15 | 96 | 10.3 | 319 | 2 | AAW88016 Mature Ps |
| 16 | 96 | 10.3 | 319 | 2 | AAW88017 Mature Ps |
| 17 | 96 | 10.3 | 319 | 2 | AAW09624 Pseudomon |
| 18 | 96 | 10.3 | 319 | 2 | AAW09625 Pseudomon |
| 19 | 96 | 10.3 | 319 | 7 | ADD13298 Pseudomon |
| 20 | 96 | 10.3 | 339 | 2 | AAW52656 E. coli Om |
| 21 | 96 | 10.3 | 358 | 2 | AAW10614 Pseudomon |
| 22 | 96 | 10.3 | 358 | 2 | AAW20177 P. glumae |
| 23 | 96 | 10.3 | 360 | 2 | AAW10438 Pseudomon |
| 24 | 96 | 10.3 | 361 | 2 | AAW10437 Pseudomon |
| 25 | 96 | 10.3 | 631 | 4 | ABG27846 Novel hum |

| | | | | | |
|----|------|------|-----|---|--------------------|
| 26 | 94 | 10.0 | 319 | 2 | AAW88010 Mature Ps |
| 27 | 93.5 | 10.0 | 319 | 2 | AAW88012 Mature Ps |
| 28 | 93.5 | 10.0 | 319 | 2 | AAW88011 Mature Ps |
| 29 | 93.5 | 10.0 | 364 | 2 | AAW77245 P. cepaci |
| 30 | 92 | 9.8 | 289 | 2 | AAW53933 Pseudomon |
| 31 | 91.5 | 9.8 | 262 | 2 | AAW23073 Thermococ |
| 32 | 91.5 | 9.8 | 262 | 6 | ABU56926 Thermococ |
| 33 | 91.5 | 9.8 | 262 | 7 | ABG75902 Thermococ |
| 34 | 91.5 | 9.8 | 262 | 7 | AAW38797 Thermococ |
| 35 | 91 | 9.7 | 313 | 2 | AAW85077 Pseudomon |
| 36 | 91 | 9.7 | 320 | 2 | AAW79118 Pseudomon |
| 37 | 91 | 9.7 | 320 | 2 | AAW09620 Pseudomon |
| 38 | 91 | 9.7 | 320 | 3 | AAW45743 Arabidops |
| 39 | 91 | 9.7 | 336 | 3 | AAW45742 Arabidops |
| 40 | 91 | 9.7 | 337 | 3 | AAW45741 Arabidops |
| 41 | 91 | 9.7 | 364 | 2 | AAW34633 Lipase B |
| 42 | 90.5 | 9.7 | 320 | 2 | AAW09622 Pseudomon |
| 43 | 90.5 | 9.7 | 363 | 2 | AAW61276 Burkholde |
| 44 | 90.5 | 9.7 | 363 | 2 | AAW6101 E. coli e |
| 45 | 90.5 | 9.7 | 363 | 4 | AAW73542 Burkholde |

ALIGNMENTS

| | | |
|----------|--|----------------------------|
| RESULT 1 | AAW23078 | standard; protein; 201 AA. |
| ID | AAW23078 | standard; protein; 201 AA. |
| XX | AAW23078; | |
| AC | AAW23078; | |
| DT | 27-AUG-2003 (revised) | |
| DT | 17-FEB-1998 (first entry) | |
| XX | XX | |
| DE | LA11.1 esterase es2. | |
| XX | XX | |
| KW | Esterase; thermostable enzyme; ester; chiral compound; cheese; pulp; | |
| KW | paper; lignin removal; sugar; lignocellulose; disease resistance; | |
| KW | feedstuff. | |
| XX | XX | |
| OS | Unidentified. | |
| XX | XX | |
| PN | WO9730160-A1. | |
| XX | XX | |
| PD | 21-AUG-1997. | |
| XX | XX | |
| PF | 11-FEB-1997; 97WO-US002039. | |
| XX | XX | |
| PR | 16-FEB-1996; 96US-00602359. | |
| XX | XX | |
| PA | (RECO-) RECOMBINANT BIOCATALYSIS INC. | |
| XX | XX | |
| PI | Robertson DE, Murphy D, Reid J, Maffia AM, Link S, Swanson RV; | |
| PI | Warren PV, Kosmicka A, Callen W, | |
| XX | XX | |
| DR | WPI, 1997-425035/39. | |
| DR | N-PSDB; AAT79331. | |
| XX | XX | |
| PT | Nucleic acid encoding heat stable esterase from thermophilic bacteria - | |
| PT | which is active in organic solvents, useful in cheese or paper | |
| PT | manufacture, and to study plant resistance to disease. | |
| XX | XX | |
| PS | Disclosure; Page 63-64; 113pp; English. | |
| XX | XX | |
| CC | This protein comprises LA11.1 esterase es2. Newly identified | |
| CC | polynucleotides (AAT79321-40) encoding esterases (AAW23069-88), some of | |
| CC | which are claimed, can be used for recombinant production of the enzymes | |
| CC | in host cells. The esterases are stable at high temperature and in | |
| CC | organic solvents, making them superior for use in production of pure | |
| CC | chiral compounds used in pharmaceutical, agricultural and other chemical | |
| CC | industries. A method is claimed for transferring an amino group from an | |
| CC | amino acid to an alpha-keto acid using a claimed esterase. The enzymes | |
| CC | may also be useful as ripening starters in cheese making, in lignin | |


```

PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148311P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155138P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156596P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157755P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158233P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

```

```

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

```

```

Query Match 10.5%; Score 98; DB 3; Length 323;
Best Local Similarity 24.0%; Pred. No. 0.05;
Matches 41; Conservativity 18; Mismatches 44; Indels 68; Gaps 7;

```

```

QY 3 IIVHGLY-----MGLVMEPLSHRLHKGYRTQTSYNSLAIDDEAFRLRLSLTH 55
Db 55 LVFVHSTHAACWAEHLPPRSS-----GFDYVANSILQGSDEPLGTVAETLQTH 108
QY 56 A-----SPNALVGHSLGLVYIKRYLESRAPSCETLSHVVAIGSPLOQASIVN 102
Db 109 ASDIADFIRESNLGSSPPVLIGHSPGLIYQYLL-----ANIVN 146
QY 103 KIEQLGLVVALGNSAEFLGKHHDDSRYPQKGS--IAGTIPLGRLSLLR 151
Db 147 K-QSLG-----TENAPPELGAVMGCVSPPSGNSGLVLR 179

```

RESULT 3

ID MAG20857 stranded, protein, 339 AA.

XX MAG20857;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23202.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridization assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131448P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132484P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138547P.
PR 10-JUN-1999; 99US-0138647P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143424P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144326P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 19-JUL-1999; 99US-0144336P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145219P.
PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145265P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148119P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148665P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149265P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150665P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 03-OCT-1999; 99US-0157117P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 23-OCT-1999; 99US-0160981P.
PR 23-OCT-1999; 99US-0160989P.

PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161358P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 10.5%; Score 98; DB 3; Length 339;
Best Local Similarity 24.0%; Pred. No. 0.053;
Matches 41; Conservative 18; Mismatches 44; Indels 68; Gaps 7;

QY 3 ILVHGLY-----MGLVHNPISHRHKLGYRTQINYSVLALIDEAIFRRDLRLTH 55
DB 71 LVFVHGSYHAMCMNHEWLPFFSSS-----GFDYSYAVSLGGESEDEPIGTAGTQTH 124
QY 56 A-----SPNALVGHSLGLVYIKRYLESRAPSCETLSHVVAIGSPLOGASIVN 102
DB 125 ASDIADFIESNLGSSPPYLIGHSPGLIVQYLL-----ANTVN 162
QY 103 KIEQLGLGVALGNSEFGLKEHDESKRYPKSGS--TAGTTPUGLRSLLR 151
DB 163 K-QSLG-----TENAFPELSGAVMVCVPPSGNSGLVLR 195

RESULT 4

AA020856
ID AA020856 standard; protein, 340 AA.

XX AA020856;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23201.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126765P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134841P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135533P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139534P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 19-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 06-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 09-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148177P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.

PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 28-OCT-1999; 99US-0162142P.

Query Match 10.5%; Score 98; DB 3; Length 340;
 Best Local Similarity 24.0%; Pred. No. 0.053;
 Matches 41; Conservative 18; Mismatches 44; Indels 68; Gaps 7;

QY 3 IILVGLY-----MHGLVWEPPLSHRLKCYRQTSYNLAIDDEAFRLRLSLTH 55
 DB 72 LVFVHGSYHAWCMWAEHMLPFSSS-----GFDYAVSLDQGESDEPLGIVAGTLOTH 125
 QY 56 A-----SPNALVGHSLGLGVYKRYLESRAPSCETLSHVVAIGSPLQASIVN 102
 DB 126 ASDIADPIESNLGSPVLIGHSPGLIYOYLL-----ANIVN 163
 QY 103 KTEQLGLVALGNSAEFLKHHDDSRYPKSGS--IAGTIPLGRLSLR 151
 DB 164 K-QSLG-----TENAFELSGAVWVGSVPSPSGSLVLR 196

RESULT 5

ID AAP90065 standard; protein; 277 AA.

XX AAP90065;

DT 25-MAR-2003 (revised)
 DT 01-NOV-1989 (first entry)

DE Sequence encoding a Pseudomonas lipase.

XX Lipase; Pseudomonas; Pseudomonas fragi.

OS Pseudomonas fragi.

PN EP318775-A.

XX 07-JUN-1989.

PF 18-NOV-1988; 88EP-00119211.

PR 03-DEC-1987; 87JP-00306638.

PR 20-MAY-1988; 88JP-00123672.

PR 27-JUL-1988; 88JP-00187684.

PA (CHCC) CHISSO CORP.

PI Aoyama S, Inouye S, Yoshida N;

XX WPI; 1989-166833/23.

PT DNA coding for a lipase - obtd. by digestion of Pseudomonas genus

XX bacterium DNA and transformation of E. coli cells.

PS Claim 1; Fig 4; 15p; English.

CC Amino acid sequence encoding a Pseudomonas lipase. Used as a lipid-

CC hydrolysing enzyme for oils and fats processing; in medicine, diagnosis,

CC in detergents and digestants. It can catalyse ester hydrolysis, ester

CC synthesis or ester conversion, in the prodn. of chemical prods. esp.

CC optically active cpds. The DNA can be obtd. from P. fragi and cloned and

XX expressed in E. coli. (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 277 AA;

| | | | | |
|-----------------------|-------|----------------|------|---------------|
| Query Match | 10.4% | Score 97 | DB 1 | Length 277 |
| Best Local Similarity | 24.6% | Pred No. 0.051 | 79 | Indels 46 |
| Matches | 51 | Conservative | 31 | Mismatches 11 |

| | | | |
|----|-----|---|-----|
| QY | 3 | ILVLVGL-----YMGGLVWAPLHRLHKLGYRTQTISYNSLAIDEA---IFRR | 48 |
| DB | 11 | ILVLVGLGFGRDRISGHHYFHG-----IKQALNECG-ASVFPVPIISAUNDNEARGDQLKQ | 64 |
| QY | 49 | LD--RSLTHASPNALVHGSLGSLVTKRYLSEAPSCETLSHVAALGSLQGASLYNKLE | 105 |
| DB | 65 | IHNTRQVGAQRVNIHGHSQALT-ARVVAALAP--ELIASVTSVSGNHSGSELADRLRL | 121 |
| QY | 106 | -----QLGLGVALGNSAEFG--LKEHDEDSRYPKSGSIAGTIPPLGLRSLILRDLPSDG | 158 |
| DB | 122 | AFVPRGLGETVAAALUTFSFSAFLSLGSHPRLPQVA-----LVAL---NALTTDG | 168 |
| QY | 159 | TVVYETKIAGTTHIAISTSYENAV | 185 |
| DB | 169 | VAAFNROYPGGLPDRWGGMGPQVAV | 195 |

RESULT 6

AA48085
ID AA48085 standard; protein; 277 AA.

| | | |
|----|-------------|---------------|
| DT | 25-MAR-2003 | (revised) |
| DT | 19-JUL-1994 | (first entry) |

DE P. fragi lipase A.

KN Lipase A; thermostability; specific activity; mechanical pulp;
 KN ester hydrolysis; ester synthesis; interesterification; acid pH; pitch;
 KN paper-making; detergent; digestive enzyme; cystic fibrosis.

OS *Pseudomonas fragi*.

PN W09401541-A1

PD 20-JAN-1994.

PF 05-JUL-1993; 93WO-DK000225.

PR 06-JUL-1992; 92DK-00000888.

PA (NOVO) NOVO-NORDISK AS.

PI Svendsen A, Pathar S, Egel-Mitani M, Borch K, Clausen G;

XX

2000

PT substitutions to increase specific activity against substrates compared

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

CC This sequence represents the P. Fungi lipase protein. This lipase has
CC high thermostability and is active at acid pH. A variant of the mature
CC lipase A in which Phe23 has been replaced by a Trp residue has increased
CC specific activity compared to the parent lipase. The lipase or variant
CC may be used in ester hydrolysis, ester synthesis or interesterification.
CC They can also be used for avoiding pitch trouble in the process for the
CC production of mechanical pulp or a paper-making process using mechanical
CC pulp. They can also be used in detergents or as a digestive enzyme, eg,
CC in the treatment of cystic fibrosis. (Updated on 25-MAR-2003 to correct
CC Pn field.)

SQ Sequence 277 AA;

| | | | | | |
|----|--|-------|------------------|-------|------------------------------------|
| | Query Match | 10.4% | Score 97; | DB 2; | Length 277; |
| | Best Local Similarity | 24.6% | Pred. No. 0.051; | | |
| | Matches | 51; | Conservative | 31; | Mismatches 79; Indels 46; Gaps 11. |
| Qy | 3 ILLVHGL-----YMGVLVMPPLSHRLHLGYRTQTITSYNSLAIDDEA---IFRR | 48 | | | |
| Dd | 11 ILLVHGFGFGRIGSHHYFFHG----IKQALNECG-ASVFVPITISAANDNEARGDQLKQ | 64 | | | |
| Qy | 49 LD--RSLTHASPNAJGHSJLGLVIKYRYLESRAESCETLSHVVAIGSPLOASTYVKIE- | 105 | | | |
| Dd | 65 IHNTIRVOYGARVNLIGHSGCALT-ARYVAIAIP-EIIASVTISGPNHSELDRIURL | 121 | | | |
| Qy | 106 -----QJGLGVYALGNSAFEG--LKHHDDSKRPYKSGSIAGTIPLGLRSLLLRPLDSG | 158 | | | |
| Dd | 122 AFVGPRIGETVAALATTSFSFAFLSALSGHPRLPQNA-----INVL---NALTTDG | 168 | | | |
| Qy | 159 TVIYEETKIAMTDHIAISTTSYENAV | 185 | | | |
| Dd | 169 VAAFNROYPOGLPDWGGMGPAQNNAV | 195 | | | |

RESULT 7

ADA33782
ID ADA33782 standard; protein; 453 AA

DT 20-NOV-2003 (first entry)

DE Acinetobacter baumannii protein #943.

KW *Acinetobacter baumannii*; bacterial disease; antibacterial; vaccine;OS *Acinetobacter baumannii*.

PN US6562958-B1

PD 13-MAY-2003.

PF 04-JUN-1999; 99US-00328352.

PR 09-JUN-1998; 98US-0088701P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Breton G, Bush D;

DR WPI; 2003-576092/54.

☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6
☐ 7
☐ 8
☐ 9
☐ 10
☐ 11
☐ 12
☐ 13
☐ 14
☐ 15
☐ 16
☐ 17
☐ 18
☐ 19
☐ 20
☐ 21
☐ 22
☐ 23
☐ 24
☐ 25
☐ 26
☐ 27
☐ 28
☐ 29
☐ 30
☐ 31
☐ 32
☐ 33
☐ 34
☐ 35
☐ 36
☐ 37
☐ 38
☐ 39
☐ 40
☐ 41
☐ 42
☐ 43
☐ 44
☐ 45
☐ 46
☐ 47
☐ 48
☐ 49
☐ 50
☐ 51
☐ 52
☐ 53
☐ 54
☐ 55
☐ 56
☐ 57
☐ 58
☐ 59
☐ 60
☐ 61
☐ 62
☐ 63
☐ 64
☐ 65
☐ 66
☐ 67
☐ 68
☐ 69
☐ 70
☐ 71
☐ 72
☐ 73
☐ 74
☐ 75
☐ 76
☐ 77
☐ 78
☐ 79
☐ 80
☐ 81
☐ 82
☐ 83
☐ 84
☐ 85
☐ 86
☐ 87
☐ 88
☐ 89
☐ 90
☐ 91
☐ 92
☐ 93
☐ 94
☐ 95
☐ 96
☐ 97
☐ 98
☐ 99
☐ 100
☐ 101
☐ 102
☐ 103
☐ 104
☐ 105
☐ 106
☐ 107
☐ 108
☐ 109
☐ 110
☐ 111
☐ 112
☐ 113
☐ 114
☐ 115
☐ 116
☐ 117
☐ 118
☐ 119
☐ 120
☐ 121
☐ 122
☐ 123
☐ 124
☐ 125
☐ 126
☐ 127
☐ 128
☐ 129
☐ 130
☐ 131
☐ 132
☐ 133
☐ 134
☐ 135
☐ 136
☐ 137
☐ 138
☐ 139
☐ 140
☐ 141
☐ 142
☐ 143
☐ 144
☐ 145
☐ 146
☐ 147
☐ 148
☐ 149
☐ 150
☐ 151
☐ 152
☐ 153
☐ 154
☐ 155
☐ 156
☐ 157
☐ 158
☐ 159
☐ 160
☐ 161
☐ 162
☐ 163
☐ 164
☐ 165
☐ 166
☐ 167
☐ 168
☐ 169
☐ 170
☐ 171
☐ 172
☐ 173
☐ 174
☐ 175
☐ 176
☐ 177
☐ 178
☐ 179
☐ 180
☐ 181
☐ 182
☐ 183
☐ 184
☐ 185
☐ 186
☐ 187
☐ 188
☐ 189
☐ 190
☐ 191
☐ 192
☐ 193
☐ 194
☐ 195
☐ 196
☐ 197
☐ 198
☐ 199
☐ 200
☐ 201
☐ 202
☐ 203
☐ 204
☐ 205
☐ 206
☐ 207
☐ 208
☐ 209
☐ 210
☐ 211
☐ 212
☐ 213
☐ 214
☐ 215
☐ 216
☐ 217
☐ 218
☐ 219
☐ 220
☐ 221
☐ 222
☐ 223
☐ 224
☐ 225
☐ 226
☐ 227
☐ 228
☐ 229
☐ 230
☐ 231
☐ 232
☐ 233
☐ 234
☐ 235
☐ 236
☐ 237
☐ 238
☐ 239
☐ 240
☐ 241
☐ 242
☐ 243
☐ 244
☐ 245
☐ 246
☐ 247
☐ 248
☐ 249
☐ 250
☐ 251
☐ 252
☐ 253
☐ 254
☐ 255
☐ 256
☐ 257
☐ 258
☐ 259
☐ 260
☐ 261
☐ 262
☐ 263
☐ 264
☐ 265
☐ 266
☐ 267
☐ 268
☐ 269
☐ 270
☐ 271
☐ 272
☐ 273
☐ 274
☐ 275
☐ 276
☐ 277
☐ 278
☐ 279
☐ 280
☐ 281
☐ 282
☐ 283
☐ 284
☐ 285
☐ 286
☐ 287
☐ 288
☐ 289
☐ 290
☐ 291
☐ 292
☐ 293
☐ 294
☐ 295
☐ 296
☐ 297
☐ 298
☐ 299
☐ 300
☐ 301
☐ 302
☐ 303
☐ 304
☐ 305
☐ 306
☐ 307
☐ 308
☐ 309
☐ 310
☐ 311
☐ 312
☐ 313
☐ 314
☐ 315
☐ 316
☐ 317
☐ 318
☐ 319
☐ 320
☐ 321
☐ 322
☐ 323
☐ 324
☐ 325
☐ 326
☐ 327
☐ 328
☐ 329
☐ 330
☐ 331
☐ 332
☐ 333
☐ 334
☐ 335
☐ 336
☐ 337
☐ 338
☐ 339
☐ 340
☐ 341
☐ 342
☐ 343
☐ 344
☐ 345
☐ 346
☐ 347
☐ 348
☐ 349
☐ 350

PT for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

Example; SEQ ID NO 5069; 3288pp; English.

CC The invention relates to isolated *Acinetobacter baumannii* nucleic acids
CC CC The *A. baumannii* nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC *A. baumannii* and other *Acinetobacter* species in a sample, in screening
CC compounds for the ability to interfere with the *A. baumannii* life cycle
CC or to inhibit *A. baumannii* infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an *A.*
CC *baumanni* protein.

SQ Sequence 453 AA;

| | | | | |
|-----------------------|-----------------|----------------|-----------|------------|
| Query Match | 10.3% | Score 96.5 | DB 6 | Length 453 |
| Best Local Similarity | 29.5% | Pred. No. 0.12 | | |
| Matches 39 | Conservative 18 | Mismatches 42 | Indels 33 | Gaps 7 |

Db 123 VQ-----DVAKTDPGLSTVIAFVNFGLVSS 152

QY 161 --TVEETKLAGMTDHIATSTSY 181

CC 153 SHNTDQDALALRLTTLTAQTATY 175

Db

RESULT 10

AA88013

AA88013 standard; protein, 319 AA.

XX

AC AA88013;

XX

DT 16-OCT-2003 (revised)

DT 02-AUG-1996 (first entry)

XX

DE Mature *Pseudomonas* glumae lipase T233R variant.

XX

KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;

KM hydrophobicity.

XX

OS *Burkholderia* glumae.

XX

FT Key Location/Qualifiers

FT Misc-difference 23

FT /note= "possible site for Phe to Arg substitution"

FT Misc-difference 129

FT /note= "possible site for Thr to Tyr substitution"

FT Misc-difference 134

FT /note= "possible site for Leu to Arg substitution"

FT Misc-difference 148

FT /note= "possible site for Thr to Val substitution"

FT Misc-difference 233

FT /label= substitution

FT /note= "Thr to Arg, in wild-type sequence a Thr residue is present at position 233 of mature *P. glumae* lipase"

FT Misc-difference 234

FT /note= "possible site for Leu to Arg substitution"

FT Misc-difference 239

FT /note= "possible site for Val to Phe substitution"

FT Misc-difference 240

FT /note= "possible site for Thr to Leu substitution"

FT Misc-difference 282

FT /note= "possible site for His to Arg substitution"

FT Misc-difference 292

FT /note= "possible site for Leu to Phe substitution"

XX

PN MO9535381-A1.

XX

PD 28-DEC-1995.

XX

PF 15-JUN-1995; 95WO-EP002349.

XX

PR 20-JUN-1994; 94EP-00201761.

XX

PA (UNIL) UNILEVER NV.

PA (UNIL) UNILEVER PLC.

XX

PI Frenken LGJ, Peters H, Suerbaum HW, De Vlieg J, Verrips CT;

XX

DR WPI; 1996-058418/06.

XX

PT *Pseudomonas* lipase variant with increased surface hydrophobicity - has

XX improved in-the-wash performance; useful in enzymatic detergent.

XX

PS Claim 12; Page; 33pp; English.

XX

CC AA88009-R88018 are protein variants of *Pseudomonas* glumae lipase. The

CC sequences are based upon the wild-type *P. glumae* sequence disclosed in

CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of

CC the amino acid substitutions indicated in the features table. Using these

CC variants it was shown that lipases can be modified in a such a way that

CC interaction with the substrate can be improved without forming large

CC hydrophobic areas on the modified lipase surface which allow aggregation

CC of lipase molecules, the lipase variants show improved in-the-wash

CC lipolytic activity and may be used in enzymatic detergent compans.

CC (updated on 16-Oct-2003 to standardise OS field)

XX

SQ Sequence 319 AA;

QY 3 IIVHGL-----YHGLVNHPLSHRLKLGRTQTISYNLAID-----EAI 45

Db 11 VILVHGLAGDKFRANVDYWGIGSDLSH-----GAXYVANLSGFGSDDDGNGRGEQL 65

QY 46 FRILDRSL--THASPRALVGHSLGGLVIRYIESRPSSETLSHYVALTSPLOQASIVVK 103

Db 66 LAVKQVLAATGATYKVNLIHSGQGILT--SRVYAAVAP--QLVAVTTTCTFRRGSEFADF 122

QY 104 IEQLGLGVALGNSAEFGLEKHDESRYPQKSGSIAGTIPLGRSLILRPDLSGTV--- 160

Db 123 VQ-----DVAKTDPGLSTVIAFVNFGLVSS 152

QY 161 --TVEETKLAGMTDHIATSTSY 181

Db 153 SHNTDQDALALRLTTLTAQTATY 175

RESULT 11

AA88015

AA88015 standard; protein, 319 AA.

XX

AC AA88015;

XX

DT 16-OCT-2003 (revised)

DT 02-AUG-1996 (first entry)

XX

DE Mature *Pseudomonas* glumae lipase V239P variant.

XX

KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;

KM hydrophobicity.

XX

OS *Burkholderia* glumae.

XX

FT Key Location/Qualifiers

FT Misc-difference 23

FT /note= "possible site for Phe to Arg substitution"

FT Misc-difference 129

FT /note= "possible site for Thr to Tyr substitution"

FT Misc-difference 134

FT /note= "possible site for Leu to Arg substitution"

FT Misc-difference 148

FT /note= "possible site for Thr to Val substitution"

FT Misc-difference 233

FT /note= "possible site for Thr to Arg substitution"

FT Misc-difference 234

FT /note= "possible site for Leu to Arg substitution"

FT Misc-difference 239

FT /label= substitution

FT /note= "Val to Phe, in wild-type sequence a Val residue is present at position 239 of mature *P. glumae* lipase"

FT Misc-difference 240

FT /note= "possible site for Thr to Leu substitution"

FT Misc-difference 282

FT /note= "possible site for His to Arg substitution"

FT Misc-difference 292

FT /note= "possible site for Leu to Phe substitution"

XX

PN MO9535381-A1.

XX

PD 28-DEC-1995.

XX

PF 15-JUN-1995; 95WO-EP002349.

XX 20-JUN-1994; 94EP-00201761.
 XX (UNITL) UNILEVER NV.
 PA (UNITL) UNILEVER PLC.
 XX
 PI Frenken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
 XX WPI; 1996-058418/06.
 DR
 XX
 PT Pseudomonas lipase variant with increased surface hydrophobicity - has
 improved in-the-wash performance, useful in enzymatic detergent.
 XX
 PS Claim 12; Page; 33pp; English.
 XX
 CC AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The
 CC sequences are based upon the wild-type P. glumae sequence disclosed in
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of
 CC the amino acid substitutions indicated in the features table. Using these
 CC variants it was shown that lipases can be modified in a such a way that
 CC interaction with the substrate can be improved without forming large
 CC hydrophobic areas on the modified lipase surface which allow aggregation
 CC of lipase molecules. The lipase variants show improved in-the-wash
 CC lipolytic activity and may be used in enzymatic detergent compans.
 CC (Updated on 16-Oct-2003 to standardise OS field)
 CC
 SQ Sequence 319 AA;
 XX
 SO
 Query Match 10.3%; Score 96; DB 2; Length 319;
 Best Local Similarity 21.2%; Pred. No. 0.081;
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;
 QY 3 IIVVHGL-----YHGLVWHPLSRLHKLGYRTQTISYNSLAID-----EAI 45
 DB 11 VILVHGLAGTDKRPANVVDYWGIGSDLOSH-----GAKYVYANLISGFQSDGDPNGRGQL 65
 QY 46 FRRLDRSL--THASPNALVGHSLGGLVTKRYLESRAVSCETLSHYVAIGSPLOGASIVNK 103
 DB 66 LAYVKQVLAATGATKTNLIGHSGGLT-SRYVAAVAP--QLVASVTTTGTIPRHGSEPADF 122
 QY 104 IEOLGVALGNSAEFLGKHEHDESRYPQKSGIAGTIPGLRSLRLRPLDSGTV-- 160
 DB 123 VQ-----DVAKTDPGTGLSSVYIAAFVAVFGTLVSS 152
 QY 161 --TVERTKIAGMTDHAISTTSY 181
 DB 153 SHNTDQDALAALRTLTAAQTATY 175
 RESULT 12
 AAR88018
 ID AAR88018 standard; protein; 319 AA.
 XX
 AC AAR88018;
 XX
 DT 16-OCT-2003 (revised)
 DT 02-AUG-1996 (first entry)
 XX
 DS Mature Pseudomonas glumae lipase L292F variant.
 XX
 KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;
 KM hydrophobicity.
 XX
 OS Burkholderia glumae.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 23 /note= "possible site for Phe to Arg substitution"
 FT Misc-difference 129 /note= "possible site for Thr to Tyr substitution"
 FT Misc-difference 134 /note= "possible site for Leu to Arg substitution"
 FT Misc-difference 148

FT /note= "possible site for Thr to Val substitution"
 FT Misc-difference 233
 FT /note= "possible site for Thr to Arg substitution"
 FT Misc-difference 234
 FT /note= "possible site for Leu to Arg substitution"
 FT Misc-difference 239
 FT /note= "possible site for Val to Phe substitution"
 FT Misc-difference 240
 FT /note= "possible site for Thr to Leu substitution"
 FT Misc-difference 282
 FT /note= "possible site for His to Arg substitution"
 FT Misc-difference 292
 FT /label= substitution
 FT /note= "Leu to Phe, in wild-type sequence a Leu residue
 is present at position 292 of mature P. glumae lipase"
 XX
 XX W09535381-A1.
 XX
 XX 28-DEC-1995.
 XX
 XX 15-JUN-1995; 95WO-EP002349.
 XX
 XX 20-JUN-1994; 94EP-00201761.
 XX
 XX (UNITL) UNILEVER NV.
 XX (UNITL) UNILEVER PLC.
 XX
 PI Frenken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
 XX WPI; 1996-058418/06.
 DR
 XX
 PT Pseudomonas lipase variant with increased surface hydrophobicity - has
 improved in-the-wash performance, useful in enzymatic detergent.
 XX
 PS Claim 12; Page; 33pp; English.
 XX
 CC AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The
 CC sequences are based upon the wild-type P. glumae sequence disclosed in
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of
 CC the amino acid substitutions indicated in the features table. Using these
 CC variants it was shown that lipases can be modified in a such a way that
 CC interaction with the substrate can be improved without forming large
 CC hydrophobic areas on the modified lipase surface which allow aggregation
 CC of lipase molecules. The lipase variants show improved in-the-wash
 CC lipolytic activity and may be used in enzymatic detergent compans.
 CC (Updated on 16-Oct-2003 to standardise OS field)
 CC
 SQ Sequence 319 AA;
 XX
 SO
 Query Match 10.3%; Score 96; DB 2; Length 319;
 Best Local Similarity 21.2%; Pred. No. 0.081;
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;
 QY 3 IIVVHGL-----YHGLVWHPLSRLHKLGYRTQTISYNSLAID-----EAI 45
 DB 11 VILVHGLAGTDKRPANVVDYWGIGSDLOSH-----GAKYVYANLISGFQSDGDPNGRGQL 65
 QY 46 FRRLDRSL--THASPNALVGHSLGGLVTKRYLESRAVSCETLSHYVAIGSPLOGASIVNK 103
 DB 66 LAYVKQVLAATGATKTNLIGHSGGLT-SRYVAAVAP--QLVASVTTTGTIPRHGSEPADF 122
 QY 104 IEOLGVALGNSAEFLGKHEHDESRYPQKSGIAGTIPGLRSLRLRPLDSGTV-- 160
 DB 123 VQ-----DVAKTDPGTGLSSVYIAAFVAVFGTLVSS 152
 QY 161 --TVERTKIAGMTDHAISTTSY 181
 DB 153 SHNTDQDALAALRTLTAAQTATY 175
 RESULT 13
 AAR88009
 ID AAR88009 standard; protein; 319 AA.


```

XX AC AAR88009;
XX XX 16-OCT-2003 (revised)
XX DT 02-AUG-1996 (first entry)
XX XX Mature Pseudomonas glumae lipase F23R variant.
XX DE Mature Pseudomonas glumae lipase F23R variant.
XX XX Mutant; lipase; enzymatic detergent; substitution; variant; improved;
XX KM hydrophobicity.
XX OS Burkholderia glumae.
XX FH Key Location/Qualifiers
XX FT Misc-difference 23 /label= substitution
XX FT /note= "Phe to Arg, in wild-type sequence a Phe residue
XX FT is present at position 23 of mature P. glumae lipase"
XX FT Misc-difference 129 /note= "possible site for Thr to Tyr substitution"
XX FT Misc-difference 134 /note= "possible site for Leu to Arg substitution"
XX FT Misc-difference 148 /note= "possible site for Thr to Val substitution"
XX FT Misc-difference 233 /note= "possible site for Thr to Arg substitution"
XX FT Misc-difference 234 /note= "possible site for Leu to Arg substitution"
XX FT Misc-difference 239 /note= "possible site for Val to Phe substitution"
XX FT Misc-difference 240 /note= "possible site for Thr to Leu substitution"
XX FT Misc-difference 282 /note= "possible site for His to Arg substitution"
XX FT Misc-difference 292 /note= "possible site for Leu to Phe substitution"
XX FT W09535381-A1.
XX PN 28-DEC-1995.
XX PD 15-JUN-1995; 95WO-BP002349.
XX XX 20-JUN-1994; 94EP-00201761.
XX PA (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX FT Franken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
XX DR WPI; 1996-058418/06.
XX XX Pseudomonas lipase variant with increased surface hydrophobicity - has
XX FT improved in-the-wash performance, useful in enzymatic detergent.
XX PS Claim 12; Page; 33pp; English.
XX XX AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The
XX CC sequences are based upon the wild-type P. glumae sequence disclosed in
XX CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of
XX CC the amino acid substitutions indicated in the features table. Using these
XX CC variants it was shown that lipases can be modified in a such a way that
XX CC interaction with the substrate can be improved without forming large
XX CC hydrophobic areas on the modified lipase surface which allow aggregation
XX CC of lipase molecules, the lipase variants show improved in-the-wash
XX CC lipolytic activity and may be used in enzymatic detergent compans.
XX CC (Updated on 16-OCT-2003 to standardise OS field)
XX SO Sequence 319 AA;
XX Query Match 10.3%; Score 96; DB 2; Length 319;
XX Best Local Similarity 21.2%; Pred. No. 0.081;
XX Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;

```

```

QY 3 IIVYGL-----YAGLVNHPISRLAKGTRTQTSYNSLAID-----EAI 45
DB 11 VILVGLAGTDRANVVDVYWGIGSDLSH-----GAVYVANSIGFSDGDPNGRGQL 65
QY 46 FRLLDRL--THASPNALVGHSLGGLVVKRYLESRAPSCETLSHVVALGSPLOGASIVNK 103
DB 66 LAYVQVLAATATATKYNLIGHGGSLT--SRVYAAVAP--QVAVSYTTIGTPHRSSEFADF 122
QY 104 IEQLGLVVALGNSEAFGLKEHDESRYPQKSGSIAGTIPLGIRSLRLDPLDSGTV--- 160
DB 123 VQ-----DVAKTDPFGISSTVIAAFVVFGLTVSS 152
QY 161 --TVEETKIAGMTDHAISTTSY 181
DB 153 SHNTQDAAALRLTTLTAQTATY 175
RESULT 14
AAR88014
ID AAR88014 standard; protein; 319 AA.
XX AC AAR88014;
XX XX 16-OCT-2003 (revised)
XX DT 02-AUG-1996 (first entry)
XX XX Mature Pseudomonas glumae lipase L234R variant.
XX XX Mutant; lipase; enzymatic detergent; substitution; variant; improved;
XX KW hydrophobicity.
XX OS Burkholderia glumae.
XX FH Key Location/Qualifiers
XX FT Misc-difference 23 /note= "possible site for Phe to Arg substitution"
XX FT Misc-difference 129 /note= "possible site for Thr to Tyr substitution"
XX FT Misc-difference 134 /note= "possible site for Leu to Arg substitution"
XX FT Misc-difference 148 /note= "possible site for Thr to Val substitution"
XX FT Misc-difference 233 /note= "possible site for Thr to Arg substitution"
XX FT Misc-difference 234 /note= "possible site for Thr to Arg substitution"
XX FT /label= substitution
XX FT /note= "Leu to Arg, in wild-type sequence a Leu residue
XX FT is present at position 234 of mature P. glumae lipase"
XX FT Misc-difference 239 /note= "possible site for Val to Phe substitution"
XX FT Misc-difference 240 /note= "possible site for Thr to Leu substitution"
XX FT Misc-difference 282 /note= "possible site for His to Arg substitution"
XX FT Misc-difference 292 /note= "possible site for Leu to Phe substitution"
XX FT W09535381-A1.
XX PN 28-DEC-1995.
XX PD 15-JUN-1995; 95WO-BP002349.
XX XX 20-JUN-1994; 94EP-00201761.
XX PA (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX FT Franken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
XX DR WPI; 1996-058418/06.

```

PT Pseudomonas lipase variant with increased surface hydrophobicity - has
 PT improved in-the-wash performance, useful in enzymatic detergent.
 XX
 PS Claim 12; Page; 33pp; English.
 CC AAR8009-R88018 are protein variants of Pseudomonas glumae lipase. The
 CC sequences are based upon the wild-type P. glumae sequence disclosed in
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of
 CC the amino acid substitutions indicated in the features table. Using these
 CC variants it was shown that lipases can be modified in a such a way that
 CC interaction with the substrate can be improved without forming large
 CC hydrophobic areas on the modified lipase surface which allow aggregation
 CC of lipase molecules. The lipase variants show improved in-the-wash
 CC lipolytic activity and may be used in enzymatic detergent compans.
 CC (Updated on 16-OCT-2003 to standardise OS field)
 SO Sequence 319 AA;
 Query Match 10.3%; Score 96; DB 2; Length 319;
 Best Local Similarity 21.2%; Pred. No. 0.081;
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;
 QY 3 IILVHGL-----YMHGLVWHPLSRLHKLGYRQTISYNSLAIDP-----EAI 45
 DB 11 VILVHGLAGTDKRFANVVDYVYGIQSDLOSH-----GAKYVYANLSGFQSDGDPNGRGEQL 65
 QY 46 FRRLDRSL--THASPNALVGHSLGGLVIRKYLESRAPSCETLSHVAIGSPLOQASIVNK 103
 DB 66 LAVYKQVLAATGATKTNKILGHSGQGLT-SRYVAAVAP--QLVASVTTTGTPHRSSEFADF 122
 QY 104 IEOLGLGVALGNSAEFGKHEHDESRYPKSGSIAGTIPLGRLSLRLDPLDSGTV--- 160
 DB 123 VQ-----DVAKTDPGSLSTVIAAFVNVFGTLVSS 152
 QY 161 --TWERTKIAGMTDHAISTSY 181
 DB 153 SHNTDDALALRLTLTAQTATY 175
 RESULT 15
 AAR8016
 ID AAR8016 standard; protein; 319 AA.
 XX
 AC AAR8016;
 XX
 DT 16-OCT-2003 (revised)
 DT 02-AUG-1996 (first entry)
 XX
 DE Mature Pseudomonas glumae lipase T240L variant.
 XX
 KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;
 KM hydrophobicity.
 KW
 XX Burkholderia glumae.
 OS
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 23 /note= "possible site for Phe to Arg substitution"
 FT MISC-difference 129 /note= "possible site for Thr to Tyr substitution"
 FT MISC-difference 134 /note= "possible site for Thr to Arg substitution"
 FT MISC-difference 148 /note= "possible site for Leu to Arg substitution"
 FT MISC-difference 148 /note= "possible site for Thr to Val substitution"
 FT MISC-difference 233 /note= "possible site for Thr to Arg substitution"
 FT MISC-difference 234 /note= "possible site for Thr to Arg substitution"
 FT MISC-difference 239 /note= "possible site for Leu to Arg substitution"
 FT MISC-difference 240 /note= "possible site for Val to Phe substitution"
 FT /label= substitution
 FT /note= "Thr to Leu, in wild-type sequence a Thr residue"

FT is present at position 240 of mature P. glumae lipase"
 FT MISC-difference 282 /note= "possible site for His to Arg substitution"
 FT MISC-difference 292 /note= "possible site for Leu to Phe substitution"
 FT
 XX WO9535381-A1.
 XX
 XX 28-DEC-1995.
 XX
 XX 15-JUN-1995; 95WO-EP002349.
 XX
 XX 20-JUN-1994; 94EP-00201761.
 XX
 XX (UNIL) UNILEVER NV.
 XX (UNIL) UNILEVER PLC.
 XX
 XX Frenken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
 XX WPI; 1996-058418/06.
 DR
 XX Pseudomonas lipase variant with increased surface hydrophobicity - has
 PT improved in-the-wash performance, useful in enzymatic detergent.
 XX
 PS Claim 12; Page; 33pp; English.
 XX
 CC AAR8009-R88018 are protein variants of Pseudomonas glumae lipase. The
 CC sequences are based upon the wild-type P. glumae sequence disclosed in
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of
 CC the amino acid substitutions indicated in the features table. Using these
 CC variants it was shown that lipases can be modified in a such a way that
 CC interaction with the substrate can be improved without forming large
 CC hydrophobic areas on the modified lipase surface which allow aggregation
 CC of lipase molecules. The lipase variants show improved in-the-wash
 CC lipolytic activity and may be used in enzymatic detergent compans.
 CC (Updated on 16-OCT-2003 to standardise OS field)
 SO Sequence 319 AA;
 Query Match 10.3%; Score 96; DB 2; Length 319;
 Best Local Similarity 21.2%; Pred. No. 0.081;
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;
 QY 3 IILVHGL-----YMHGLVWHPLSRLHKLGYRQTISYNSLAIDP-----EAI 45
 DB 11 VILVHGLAGTDKRFANVVDYVYGIQSDLOSH-----GAKYVYANLSGFQSDGDPNGRGEQL 65
 QY 46 FRRLDRSL--THASPNALVGHSLGGLVIRKYLESRAPSCETLSHVAIGSPLOQASIVNK 103
 DB 66 LAVYKQVLAATGATKTNKILGHSGQGLT-SRYVAAVAP--QLVASVTTTGTPHRSSEFADF 122
 QY 104 IEOLGLGVALGNSAEFGKHEHDESRYPKSGSIAGTIPLGRLSLRLDPLDSGTV--- 160
 DB 123 VQ-----DVAKTDPGSLSTVIAAFVNVFGTLVSS 152
 QY 161 --TWERTKIAGMTDHAISTSY 181
 DB 153 SHNTDDALALRLTLTAQTATY 175
 Search completed: April 29, 2004, 11:35:16
 Job time : 60 secs

| | | | | | | |
|----|-------|------|------|---|--------|---------------------|
| 1 | 532 | 56.8 | 209 | 2 | B82501 | hypothetical prote |
| 2 | 121.5 | 13.0 | 249 | 2 | S76547 | hypothetical prote |
| 3 | 112.5 | 12.0 | 251 | 2 | H82664 | conserved hypothet |
| 4 | 105.5 | 11.3 | 229 | 2 | AG3185 | conserved hypothet |
| 5 | 104.5 | 11.2 | 243 | 2 | C95400 | hypothetical prote |
| 6 | 103.5 | 11.1 | 202 | 2 | S77556 | hypothetical prote |
| 7 | 102 | 10.9 | 339 | 2 | C82466 | lactonizing lipase |
| 8 | 99.5 | 10.6 | 211 | 2 | AE1975 | hypothetical prote |
| 9 | 97 | 10.4 | 277 | 1 | S02005 | triacylglycerol 1 |
| 10 | 96.5 | 10.3 | 309 | 2 | G83044 | lipase hIpc PA4813 |
| 11 | 96 | 10.3 | 327 | 2 | T35482 | hypothetical prote |
| 12 | 96 | 10.3 | 358 | 1 | A48952 | triacylglycerol 1 |
| 13 | 95.5 | 10.2 | 340 | 2 | A86000 | hypothetical prote |
| 14 | 95.5 | 10.2 | 340 | 2 | D65129 | hypothetical 38.5 |
| 15 | 95.5 | 10.2 | 340 | 2 | D91154 | hypothetical prote |
| 16 | 95 | 10.1 | 255 | 2 | AH3167 | conserved hypothet |
| 17 | 94 | 10.0 | 332 | 2 | F83600 | conserved hypothet |
| 18 | 92.5 | 9.9 | 284 | 2 | A82133 | conserved hypothet |
| 19 | 92 | 9.8 | 290 | 2 | T36757 | probable secreted |
| 20 | 90.5 | 9.7 | 300 | 2 | AD1551 | hypothetical prote |
| 21 | 90.5 | 9.7 | 364 | 1 | UT0579 | triacylglycerol 1 |
| 22 | 90.5 | 9.7 | 1215 | 2 | S50438 | probable Cas2+-tran |
| 23 | 87 | 9.3 | 345 | 2 | F71261 | probable lipase - |
| 24 | 86.5 | 9.2 | 237 | 2 | G75316 | probable lipase - |
| 25 | 86 | 9.2 | 372 | 2 | S75587 | H+/Ca2+ exchanging |
| 26 | 85 | 9.1 | 1216 | 2 | S46177 | probable Cas2+-tran |
| 27 | 84.5 | 9.0 | 171 | 2 | S15911 | hypothetical prote |
| 28 | 84.5 | 9.0 | 299 | 2 | AF1193 | hypothetical prote |
| 29 | 84.5 | 9.0 | 771 | 2 | H72410 | hypothetical prote |

| | | | | | | |
|----|------|-----|-------|---|--------|----------------------|
| 30 | 84 | 9.0 | 21.7 | 2 | A63426 | hypothetical prote |
| 31 | 84 | 9.0 | 24.7 | 2 | A83712 | hypothetical prote |
| 32 | 83.5 | 8.9 | 26.3 | 2 | H82134 | conserved hypothet |
| 33 | 82.5 | 8.8 | 26.3 | 2 | T01149 | probable acetone-c |
| 34 | 82.5 | 8.8 | 33.3 | 2 | B70906 | probable lysophos |
| 35 | 81.5 | 8.7 | 29.9 | 2 | D75007 | lysophospholipase |
| 36 | 81.5 | 8.7 | 33.5 | 2 | E86533 | probable leader 19 |
| 37 | 81.5 | 8.7 | 33.5 | 2 | E81578 | conserved hypothet |
| 38 | 81.5 | 8.7 | 133.3 | 2 | S33635 | blastocysta polyprot |
| 39 | 80.5 | 8.6 | 26.8 | 2 | S47087 | rib7 protein - ri |
| 40 | 80.5 | 8.6 | 59.1 | 2 | E83729 | adenine deaminase |
| 41 | 80.5 | 8.6 | 146.0 | 2 | D81675 | polymorphic membra |
| 42 | 80 | 8.5 | 26.7 | 2 | G82119 | probable esterase/ |
| 43 | 80 | 8.5 | 57.4 | 2 | T47566 | hypothetical prote |
| 44 | 79.5 | 8.5 | 1.79 | 2 | H83755 | hypothetical prote |
| 45 | 79.5 | 8.5 | 24.9 | 2 | H87675 | hydrolase, alpha/b |

ALIGNMENTS

RESULT 1

Hypothetical protein VCA0092 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
 C|Species: *Vibrio cholerae*
 C|Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C|Accession: B82501
 R|Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Claydon, R.A.; Gwinn, M.L.; Dodson, R.J.
 |charadson, D.; Ermolaeva, M.D.; Vamthavan, V.; Bass, S.; Qin, H.; Driscoll, I.; Sellers,
 |, R.R.; Melancon, J.J.; Venter, J.C.; Fraser, C.M.
 |Nature 406, 477-483, 2000
 A|Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A|Reference number: A82055; MUID:20406833; PMID:10952301
 A|Accession: B82501
 A|Status: Preliminary
 A|Molecule type: DNA
 A|Residues: 1-209 <HE>
 A|Cross-references: GB:AB004352; GB:AE003853; NID:G9657475; PIDN:AAF96006.1; GSPDB:GN000
 |A|Experimental source: serogroup O1; strain N16961; biotype El Tor
 C|GeneticS:
 A|Gene: VCA0092
 A|Map position: 2

| Query Match | Score | DB 2 | Length |
|-------------|-------|------|--------|
| 56.8% | 532 | 209 | |

Matches 104; Conservative 28; Mismatches 44; Indels 0; Gaps 0;

QY 1 MÖILVHGLVHPLSHRLHKLGYRTÖTISYNSLAIDDEAIFRDLRSLTHASPNA 60

Db 1 MKIVILHGLYMHGLVMQPLAQRLNKLGYQTEVISYNTLAIIDDEKVFQTI DSALAQDRINV 60

61 LVGHSGLVIRYLESRAPSCETLSHVVAIGSPLOGASIVNKIEQLGLGVALGNSAEFG 120

D6 61 LVGHS LGGLMIKHYLRSRHPSPNVISHTVALLASPLKGASIVPKIQQLGLGAMLGNHLYG 120

121 LKEHDESRYPQKSGSIAGTIPLGLRSLLRDPLDS DGT VTEETKIAGMTDHI AI 176

DB 121 LQLHODSWELPQRLSCIAGILRFGRPIILGGSGMCDGTVIAEIQISGMIDHLLL 1/6

RESULT 2

S/6547
hypothetical protein - Symechovatis sp. (strain PCC 6803)

C/species: *Synechocystis* sp.
A: Variety: PCC 6803

```
C:\Access\10: 23-Apr-1997 #sequence_rev18100 23-Apr-1997 #revl_change 08-Dec-1999
```

O., K.: Okumura, S.: Shimo, S.: Takeuchi, C.: Wada, T.: Watanabe, A.: Yamada,

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechococcus* sp. 105-130, 1990

A;Reference number: S74322; MUID:97061201; PMID:8905231

C:Superfamily: Pseudomonas triacylglycerol lipase
C:Keywords: carboxylic ester hydrolase; extracellular protein

Query Match 10.3%; Score 96; DB 1; Length 388;
Best Local Similarity 21.2%; Pred. No. 0.51;
Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;

QY 3 IILVHGL-----YMHGLVWHPPLSHRLHKLGYVTRTQTSYNSLAID-----EAI 45
DB 50 VILVHGLAGTQKFAVVDVYVWYGIQSDLSH-----GAKVYVANSISGFSQSDGPRGRGQL 104
QY 46 FRELDRLSIT--THASPNALVGHSLGLVIRKYLESRAPSCETLSHVVAIGSPLOGASIVNK 103
DB 105 LAAYVQVLAATGATKVNLIHSGGGLT--SRVAAVAP--QVAAVTTITGFPHRSGEPADF 161
QY 104 IEQGLGVALGNSAFGLKHEHDESRYPQKSGSIAGTPIGLRLSLLRDPLDSGTY--- 160
DB 162 VO-----DVAKTDPTGLSTVIAAFVNFGLVSS 191
QY 161 --TVERTKIAGMTDHIATSTSY 181
DB 192 SHNTQDALALRLTITTAQTATY 214

RESULT 13

hypothetical protein yheT [imported] - Escherichia coli (strain O157:H7, substrain EDL93)

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 22-Jun-2003

A:Accession: AB6000
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Jim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: AB6000

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-340 <STO>

A:Cross-references: GB:AB005174; NID:912517975; PIDN:ANG58461.1; GSPDB:GN00145; UMGP:247
A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yheT

C:Superfamily: alpha/beta hydrolase

Query Match 10.2%; Score 95.5; DB 2; Length 340;
Best Local Similarity 24.9%; Pred. No. 0.53;

Matches 53; Conservative 27; Mismatches 60; Indels 73; Gaps 13;

QY 2 QIILVHGL-----YMHGL-----VMH-----PLSHRLHKLGYRTQTSY 35
DB 74 RLIVFHGLGSLNSPYAHGLVEAAQKRWLGVMHFRGSGEP--NRMRH-----Y 123
QY 36 NSLAIDDEAIFFR--LDRSLTHASPNALVGHSLGLVIRKYLESRAPSCETLSHVVAIGSP 94
DB 124 HSGETEDASWFLRMQLREFGHA--PTAAVGYSLGNNMLACLAKKGNL-----P 171
QY 95 LOGASIVN---KIQOLGVALGNSAFGLKHEHDESRY-----PQKSGSIAGTIP 142
DB 172 VDAAVIVSAPFMLEACSYHMEKGFSSRY-----QRYLNLTLKANAAKLAAPGTLIP 223
QY 143 LGLRSL--LIRDPDSDGTYVTEETKIAGMTDHI 174
DB 224 INLAQLKSVRIREFDILIT--ARHGYADAI 253

RESULT 14

hypothetical 38.5 kD protein in ktfb-prkb intergenic region - Escherichia coli (strain K

C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 22-Jun-2003

C:Accession: D65129

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: D65129

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-340 <BLAT>
A:Cross-references: GB:AB000411; GB:U00096; NID:92367213; PIDN:AC76378.1; PID:91789752
A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yheT

C:Superfamily: alpha/beta hydrolase

Query Match 10.2%; Score 95.5; DB 2; Length 340;
Best Local Similarity 24.9%; Pred. No. 0.53;

Matches 53; Conservative 27; Mismatches 60; Indels 73; Gaps 13;

QY 2 QIILVHGL-----YMHGL-----VMH-----PLSHRLHKLGYRTQTSY 35
DB 74 RLIVFHGLGSLNSPYAHGLVEAAQKRWLGVMHFRGSGEP--NRMRH-----Y 123
QY 36 NSLAIDDEAIFFR--LDRSLTHASPNALVGHSLGLVIRKYLESRAPSCETLSHVVAIGSP 94
DB 124 HSGETEDASWFLRMQLREFGHA--PTAAVGYSLGNNMLACLAKKGNL-----P 171
QY 95 LOGASIVN---KIQOLGVALGNSAFGLKHEHDESRY-----PQKSGSIAGTIP 142
DB 172 VDAAVIVSAPFMLEACSYHMEKGFSSRY-----QRYLNLTLKANAAKLAAPGTLIP 223
QY 143 LGLRSL--LIRDPDSDGTYVTEETKIAGMTDHI 174
DB 224 INLAQLKSVRIREFDILIT--ARHGYADAI 253

RESULT 15

hypothetical protein Ecs4204 [imported] - Escherichia coli (strain O157:H7, substrain R

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 22-Jun-2003

A:Accession: D91154

R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D91154

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-340 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037627.1; PID:913363677; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: Ecs4204

C:Superfamily: alpha/beta hydrolase

Query Match 10.2%; Score 95.5; DB 2; Length 340;
Best Local Similarity 24.9%; Pred. No. 0.53;

Matches 53; Conservative 27; Mismatches 60; Indels 73; Gaps 13;

QY 2 QIILVHGL-----YMHGL-----VMH-----PLSHRLHKLGYRTQTSY 35
DB 74 RLIVFHGLGSLNSPYAHGLVEAAQKRWLGVMHFRGSGEP--NRMRH-----Y 123
QY 36 NSLAIDDEAIFFR--LDRSLTHASPNALVGHSLGLVIRKYLESRAPSCETLSHVVAIGSP 94
DB 124 HSGETEDASWFLRMQLREFGHA--PTAAVGYSLGNNMLACLAKKGNL-----P 171
QY 95 LOGASIVN---KIQOLGVALGNSAFGLKHEHDESRY-----PQKSGSIAGTIP 142
DB 172 VDAAVIVSAPFMLEACSYHMEKGFSSRY-----QRYLNLTLKANAAKLAAPGTLIP 223
QY 143 LGLRSL--LIRDPDSDGTYVTEETKIAGMTDHI 174

Mon May 3 09:19:52 2004

us-10-603-260-5.rpr

Page 6

Db 224 INLAQKSVRIRREFDDLIT--ARIHGVDAT 253

Search completed: April 29, 2004, 11:37:13
Job time : 23 secs

DR EMBL: AE004362; AAF96133.1; ALT_INT.
 DR PIR: S15911; S15911.
 DR HSSP: P22088; 3LIP.
 DR TIGR: VCA0221; .
 DR InterPro: IPR000073; A/b hydrolase.
 DR InterPro: IPR008262; Lipase AS.
 DR InterPro: IPR000379; Ser esters.
 DR Pfam: PF00561; abhydrolase; 1.
 DR PROSITE: PS00120; LIPASE_SER; 1.
 DR HydroLase; Lipid degradation; Signal, Complete proteome.
 KM SIGNAL 1 22
 FT CHAIN 23 312
 FT ACT_SITE 110 110 LACTONIZING LIPASE.
 FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 278 278 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 312 AA; 32995 MW; 96AA00A07A5686C CRC64;
 Query Match 10.9%; Score 102; DB 1; Length 312;
 Best Local Similarity 25.9%; Pred. No. 0.12;
 Matches 49; Conservative 29; Mismatches 51; Indels 60; Gaps 10;
 QY 3 ILIVHGL-----YHGLVHPLSHRLKLG---YRQTIYSNGLAIDDEAIFRRL 49
 DB 38 ILIVHGLFPGDITLAGNDYFPHGI--PQS--LITRDAQYVAQVSAVNSSEKGEQLAQQV 92
 QY 50 DR--SLTHASPNALVGHSLGLVIRKYLESRAPSCETLSHVAISPLQASIVKIKIQL 107
 DB 93 ESLANTATGAKVNLIGHSHGCPRTI-RYVASVAP--DIVASVTSIGVHKGSADVADLV-- 147
 QY 108 GLGLGVALGNSAEFGLKEHDESRYPQKSGSIAGTIPGLRSLIL-----RDPL 154
 DB 148 --GVI-----PSGVSSEQVAVGLTQGLVALIDLLSGGKAHPQDPL 185
 QY 155 DSDGTVTVE 163
 DB 186 ASLALITTE 194
 RESULT 2
 LIP_PSEPR STANDARD; PRT; 277 AA.
 AC P08658;
 DT 01-UN-1988 (Rel. 06, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).
 OS Pseudomonas fragi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 ON NCBI_TaxID=296;
 RX STRAIN=FROM N.A.
 RC STRAIN=IFO 12049;
 RC MEDLINE=69078617; PubMed=3060375;
 RA Aoyama S., Yoshida N., Inouye S.;
 RT "Cloning, sequencing and expression of the lipase gene from
 RT Pseudomonas fragi IFO-12049 in E. coli.";
 RL FEBS Lett. 242:36-40(1988).
 RN [2]
 RN SEQUENCE OF 1-118 FROM N.A.
 RP STRAIN=IFO 3458;
 RC MEDLINE=87100109; PubMed=3800995;
 RA Kugimiyu W., Otani Y., Hashimoto Y., Takagi Y.;
 RT "Molecular cloning and nucleotide sequence of the lipase gene from
 RT Pseudomonas fragi.";
 RL Biochem. Biophys. Res. Commun. 141:185-190(1986).
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
 CC fatty acid anion.
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS
 CC LIPASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: M14604; AAA25879.1; ALT_SEQ.
 DR PIR: X14033; CAA32193.1; .
 DR PIR: S02005; S02005.
 DR HSSP: Q05489; 17AH.
 DR InterPro: IPR000073; A/b hydrolase.
 DR InterPro: IPR008262; Lipase AS.
 DR InterPro: IPR000379; Ser esters.
 DR Pfam: PF00561; abhydrolase; 1.
 DR PROSITE: PS00120; LIPASE_SER; 1.
 DR HydroLase; Lipid degradation; Signal.
 KM SIGNAL 1 18
 FT CHAIN 19 277
 FT ACT_SITE 83 83 LIPASE.
 FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 260 260 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 277 AA; 29985 MW; 9A1562B072C997 CRC64;
 Query Match 10.4%; Score 97; DB 1; Length 277;
 Best Local Similarity 24.6%; Pred. No. 0.27;
 Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;
 QY 3 ILIVHGL-----YHGLVHPLSHRLKLG---YRQTIYSNGLAIDDEAIFRRL 48
 DB 11 ILIVHGLFPGDITLAGNDYFPHGI--PQS--LITRDAQYVAQVSAVNSSEKGEQLAQQV 92
 QY 49 LD--SLTHASPNALVGHSLGLVIRKYLESRAPSCETLSHVAISPLQASIVKIKIQL 105
 DB 65 IHNLRQVGAQRVNLIGHSHGCPRTI-RYVASVAP--DIVASVTSIGVHKGSADVADLV-- 147
 QY 106 -----QGLGVALGNSAEFGLKEHDESRYPQKSGSIAGTIPGLRSLIL-----RDPL 154
 DB 122 AFVPGRLGETVAALTTSPFAFLSALSGHPRCPQNA-----LNLV--NALTTDS 168
 QY 159 TVTVEETKIAGMTDHTAISTSYENAV 185
 DB 169 VAAFNRRYPGQLPDRMGCMGPAQVNAV 195
 RESULT 3
 LIP_PSEGL STANDARD; PRT; 358 AA.
 AC Q05489;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).
 OS Pseudomonas glumae, and
 OS Chromobacterium viscosum.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia.
 ON NCBI_TaxID=337; 42739;
 RX STRAIN=FROM N.A.; SEQUENCE OF 40-61, AND MUTAGENESIS.
 RC SPECIES=P. glumae; STRAIN=PGL / CBS 322.89;
 RC MEDLINE=93119130; PubMed=1476423;
 RA Frenken L.G.J., Edmond M.R., Batendurg A.M., Bos J.W., Visser C.,
 RA Verrips C.T.;
 RT "Cloning of the Pseudomonas glumae lipase gene and determination of
 RT the active site residues.";
 RL Appl. Environ. Microbiol. 58:3787-3791(1992).
 RN [2]
 RN SEQUENCE OF 40-54, AND CHARACTERIZATION.
 RP SPECIES=C. viscosum;
 RC MEDLINE=85306500; PubMed=7786905;
 RX Taira W.A., Liebeton K., Costa J.V., Cabral J.M.S., Jaeger K.-E.;
 RA "Lipase from Chromobacterium viscosum: biochemical characterization

RT indicating homology to the lipase from *Pseudomonas glumae*.
 RL Biochim. Biophys. Acta 1256:396-402(1995).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RC SPECIES=P. glumae;
 RX MEDLINE=94009622; PubMed=8405390;
 RA Noble M.E.M., Cleasby A., Johnson L.N., Edmond M.R., Frenken L.G.J.;
 RT "The crystal structure of triacylglycerol lipase from *Pseudomonas*
 RL *glumae* reveals a partially redundant catalytic aspartate.";
 FERS Lett. 331:123-128(1993).
 [4]
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RC SPECIES=C. viscosum; STRAIN=ATCC 6918;
 RX MEDLINE=96275656; PubMed=8683577;
 RA Lang D., Hofmann B., Haalok L., Hecht H.-J., Spener F., Schmid R.D.,
 RA Schomburg D.;
 RT "Crystal structure of a bacterial lipase from *Chromobacterium*
 RL *viscosum* ATCC 6918 refined at 1.6-A resolution.";
 J. Mol. Biol. 259:704-717(1996).
 CC -1- FUNCTION: HYDROLYSIS OF TRIGLYCERIDES.
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
 CC fatty acid anion.
 CC -1- COFACTOR: Requires calcium.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS
 CC LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR EMBL: X70354; CAA49812.1; -;
 DR EMBL: A16323; CAA01279.1; -;
 DR EMBL: A32021; CAA02073.1; -;
 DR PIR: A48952; A48952.
 DR PDB: 1TAN; 31-MAY-94.
 DR PDB: 1CVL; 01-APR-97.
 DR PDB: 1QGE; 06-MAY-99.
 DR InterPro: IPR000073; A/b_hydrolase.
 DR InterPro: IPR008262; Lipase_AS.
 DR InterPro: IPR00379; Ser_estrs.
 DR Pfam: PF00561; abhydrolase; 1.
 DR ProSite: PS00120; LIPASE_SER; 1.
 KW Hydrolase; Lipid degradation; Signal; Calcium; 3D-structure.
 FT SIGNAL 1 39
 FT CHAIN 40 358
 FT ACT_SITE 126 126
 FT ACT_SITE 302 302
 FT ACT_SITE 324 324
 FT DISULFD 229 308
 FT MUTAGEN 54 54
 FT MUTAGEN 126 126
 FT MUTAGEN 160 160
 FT MUTAGEN 160 160
 FT MUTAGEN 280 280
 FT MUTAGEN 280 280
 FT MUTAGEN 302 302
 FT MUTAGEN 302 302
 FT MUTAGEN 324 324
 FT CONFLICT 40 40
 FT TURN 43 44
 FT STRAND 50 53
 FT TURN 61 62
 FT HELIX 65 67
 FT TURN 70 71
 FT HELIX 72 78
 FT TURN 79 80
 FT STRAND 83 86

FT TURN 89 90
 FT TURN 97 98
 FT HELIX 100 115
 FT TURN 116 116
 FT STRAND 120 125
 FT TURN 126 126
 FT HELIX 127 138
 FT HELIX 140 142
 FT STRAND 143 149
 FT TURN 153 154
 FT HELIX 157 168
 FT TURN 170 171
 FT TURN 173 174
 FT HELIX 176 188
 FT TURN 191 193
 FT HELIX 196 204
 FT TURN 205 206
 FT HELIX 208 217
 FT TURN 221 222
 FT STRAND 223 223
 FT TURN 231 231
 FT STRAND 231 231
 FT STRAND 235 238
 FT TURN 239 240
 FT STRAND 241 248
 FT STRAND 250 250
 FT STRAND 253 259
 FT TURN 260 261
 FT STRAND 262 267
 FT TURN 275 275
 FT HELIX 276 279
 FT HELIX 281 294
 FT TURN 295 297
 FT STRAND 301 301
 FT STRAND 306 306
 FT HELIX 307 310
 FT STRAND 314 315
 FT STRAND 320 320
 FT TURN 324 325
 FT HELIX 326 328
 FT TURN 329 333
 FT TURN 337 338
 FT HELIX 342 356
 FT TURN 357 358
 SQ SEQUENCE 358 AA; 36928 MW; FE7B5D7A22EC6B4B CRC64;
 Query Match 10.3%; Score 96; DB 1; Length 358;
 Best Local Similarity 21.2%; Pred. No. 0.45;
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;
 QY 3 IIVHGL-----YHGLVHPLSHRLHLGYRTQITISNSLAID-----EAI 45
 DB 50 VILVHGLAGTDKPFANVVVDYWGIGSLDQSH-----GAKVYVANSFGSDGDPNGRGQL 104
 QY 46 FRLDRSL--THASPNALVHSLGLVTKRYLESAPACETLSHVAIGSPLOGASIVNK 103
 DB 105 IAYVQVLAATCATNVLNIGSGSLT-SRYVAAP--QVAVSTTIGTHRGSEFFDF 161
 QY 104 IEQLGLVALGNSAEFGKHEHDESRYPQKSGISAGTIPILGRLSLRDLPSDGTV-- 160
 DB 162 VQ-----DVAKTPTGISSTVIAFAVNFGLTVSS 191
 QY 161 --TVEETKIAGMTDHIATISYSY 181
 DB 192 SHNTDDALALRITLTIAQTATY 214
 RESULT 4
 YHET ECOLI
 ID YHET ECOLI STANDARD; PRT; 340 AA.
 AC P45524;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN HYPERHET protein yhet.
 OS YHEH OR B3353.
 CC Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxId=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474 (1997).
 CC -1- SIMILARITY: Belongs to the UPF0017 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U18997; AAA58150.1; -.
 DR EMBL; AE000411; AAC76376.1; -.
 DR PIR; D65129; D65129.
 DR EcoGene; EGI2904; yhet.
 DR InterPro; IPR000073; A/b_hydrolase.
 DR InterPro; IPR000379; Ser_ester.
 DR InterPro; IPR000952; UPF0017.
 DR Pfam; PF00561; abhydrolase; 1.
 DR PROSITE; PS01133; UPF0017; 1.
 DR Hypothetical protein Transmembrane; Complete proteome.
 KM POTENTIAL.
 KW TRANSMEM 147 164
 SQ SEQUENCE 340 AA; 38495 MW; 7DB6BA6795CBA64D CRC64;
 Query Match 10.2%; Score 95.5; DB 1; Length 340;
 Best Local Similarity 24.9%; Pred. No. 0.46;
 Matches 53; Conservative 27; Mismatches 60; Indels 73; Gaps 13;
 QY 2 QILVHGL-----YMHGL-----VME-----PLSHRLKGYRTQTTISY 35
 DB 74 RLAVFEGLEGLSINSFYAHGLVAAQKRGWLVGMHFRGSGSEP--NRHRI-----Y 123
 QY 36 NSLAIDDEAIPRR-LDRSLTHASPNALVGHSLGIVIKRYLESRAVSCETLSHVVALGSP 94
 DB 124 HSGETEDASWFLRMLOERGH-PTNAVGYSLGNNMLACLAKENDL-----P 171
 QY 95 LOGASIVN--KIEOLGVALGNSAEFLKEHDESDRY-----PQSGSIAGTIP 142
 DB 172 VDAAVIVSAPFMLEASVYHEKGFSPRVY-----QRYTLNLKANAARLAAYPGTLP 223
 QY 143 LGHRSI-LRDPDSDGTYVETKTAGTIDHI 174
 DB 224 INLAQKSVRIREFDILIT---ARHGYADAI 253
 RESULT 5
 LIP_PSSS STANDARD; PRT; 364 AA.
 AC P25275;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).
 GN LIP.
 OS Pseudomonas sp. (strain KMT-56).
 OC Bacteria; Proteobacteria.
 OX NCBI_TaxId=311;

RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 45-47.
 RX MEDLINE=92118328; PubMed=1358739;
 RA Iizumi T., Nakamura K., Shimada Y., Sugihara A., Tomioka Y.,
 RA Fukase T.;
 RT "Cloning, nucleotide sequencing, and expression in Escherichia coli
 RT of a lipase and its activator genes from Pseudomonas sp. KMT-56";
 RL Agric. Biol. Chem. 55:2349-2357 (1991).
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H₂O = diacylglycerol + a
 CC fatty acid anion.
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS
 CC LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10069; BA000960.1; -.
 DR EMBL; S77842; AAC60400.1; -.
 DR HSP; P22088; 3LIP.
 DR InterPro; IPR000073; A/b_hydrolase.
 DR InterPro; IPR008262; Lipase_AS.
 DR InterPro; IPR000379; Ser_ester.
 DR Pfam; PF00561; abhydrolase; 1.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 DR HydroLase; Lipid degradation; Signal.
 KM SIGNAL 1 44
 FT CHAIN 45 364
 FT ACT_SITE 131 131
 FT ACT_SITE 286 286
 FT ACT_SITE 330 330
 SQ SEQUENCE 364 AA; 37511 MW; F346CB8B2B94E27D CRC64;
 Query Match 9.7%; Score 90.5; DB 1; Length 364;
 Best Local Similarity 26.6%; Pred. No. 1.4;
 Matches 33; Conservative 21; Mismatches 43; Indels 27; Gaps 5;
 QY 3 IILVHGL-----YMHGLVMEPLSHRLKGYRTQTTISYNSLAIDDEA----- 44
 DB 55 IILVHGLSGDKKAGVETWYG-----IQEDLQGNATVYVNLSEFGDDGANGREGOL 109
 QY 45 -IFRLDRSLTHASPNALVGHSLGIVIKRYLESRAVSCETLSHVVALGSPLOGASIVNK 103
 DB 110 LAVVKTVALATGATKVLVGHSGSGGLT--GRYVAAP--DLVASVTITGTPHRSSEFADF 166
 QY 104 IEOL 107
 DB 167 VQNV 170
 RESULT 6
 ATC6_YEAST STANDARD; PRT; 1215 AA.
 AC P39985;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable cation-transporting ATPase 1 (EC 3.6.3.-).
 GN SP1 OR YEL031W.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313264; PubMed=9169868;
 RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
 RA Araujo R., Aviles E., Bernan A., Brennan T., Carpenter J., Chen E.,

RA Cheriy J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
 RA Huncke-Smith S., Hyman R.W., Kayser A., Komp C., Laakkari D., Lew H.,
 RA Lin D., Mosedale D., Nakanahara K., Nanth A., Norgren R., Oefner P.,
 RA Oh C., Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
 RA Smith V., Taylor P., Wei Y., Weinstein D., Davis R.W.,
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.",
 RN Nature 387:78-81(1997).
 CC [2]
 CC CHARACTERIZATION.
 CC MEDLINE=99291048; PubMed=10361284;
 RA Suzuki C., Shima Y.I.;
 RT "P-type ATPase spf1 mutants show a novel resistance mechanism for the
 RT killer toxin SMK1".
 CC Mol. Microbiol. 32:813-823(1999).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily V.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U18530; AAB64508.1; -.
 CC DR PIR; S50428; S50428.
 CC DR Germonline; 139035; -.
 CC DR SGD; S0000757; SPFL.
 CC DR GO; GO:0005789; C:cytoplasmic reticulum membrane; IDA.
 CC DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IDA.
 CC DR GO; GO:0006874; P:calcium ion homeostasis; IMP.
 CC DR GO; GO:0006486; P:protein amino acid glycosylation; IMP.
 CC DR InterPro; IPR001757; ATPase_E1-E2.
 CC DR InterPro; IPR008250; E1-E2_ATPase_reg.
 CC DR InterPro; IPR005834; Hydrolyase.
 CC DR InterPro; IPR005444; P-ATPase-V.
 CC DR Pfam; PF00122; E1-E2_ATPase; 1.
 CC DR Pfam; PF00702; Hydrolyase; 1.
 CC DR PRINTS; PRO0119; CATATPASE.
 CC DR TIGRFS; TIGR01494; ATPase_P-type; 6.
 CC DR TIGRFS; TIGR01657; P-ATPase-V; 1.
 CC DR PROSITE; PS00554; ATPase_E1-E2; 1.
 CC KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
 CC FT DOMAIN 1 21
 CC FT TRANSSEM 22 43
 CC FT TRANSSEM 44 49
 CC FT TRANSSEM 50 72
 CC FT TRANSSEM 73 191
 CC FT TRANSSEM 192 214
 CC FT TRANSSEM 215 217
 CC FT TRANSSEM 218 236
 CC FT TRANSSEM 237 399
 CC FT TRANSSEM 400 419
 CC FT TRANSSEM 420 432
 CC FT TRANSSEM 433 454
 CC FT TRANSSEM 455 493
 CC FT TRANSSEM 994 1013
 CC FT TRANSSEM 1014 1020
 CC FT TRANSSEM 1021 1037
 CC FT TRANSSEM 1038 1055
 CC FT TRANSSEM 1056 1079
 CC FT TRANSSEM 1080 1099
 CC FT TRANSSEM 1100 1122
 CC FT TRANSSEM 1123 1133
 CC FT TRANSSEM 1134 1153
 CC FT TRANSSEM 1154 1170
 CC FT TRANSSEM 1171 1193
 CC FT TRANSSEM 1194 1215
 CC FT MOD RES 487 487
 CC FT METAL 816 816

FT METAL 820 820 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 1215 AA; 135268 MW; 7A960D34B91B5AE CRC64;
 Query Match 9.7%; Score 90.5; DB 1; Length 1215;
 Best Local Similarity 23.5%; Pred. No. 5.3;
 Matches 40; Conservative 22; Mismatches 53; Indels 55; Gaps 4;
 QY 41 DDEAFRRDLRSITLTHSPNALVGHSLGGIVIRKRYESRAPSCEITLTHVALGSPLOGAS 99
 DB 753 DHSKLFDRDIANTVGYALNALBSHSQLRLDLHNTWYAV-----SPQCKRF 799
 QY 100 IYKIKQLG-----LGVALGNSAFGLKEHDESR----- 129
 DB 800 LNTLTMDGVQYTLMDGDTNDVGAALKQAHVGIATLNGTEGKKGCEGRLEGKMKMYIK 859
 QY 130 -----YQKSGSIAGTIPGLRSLRLRDPDPSGCTVTEETKIA 168
 DB 860 QTEPMARKNPQPPVPEPIAHLPPEQPKPHYLKALSKGIVITPEIRKA 909
 RESULT 7
 LIP_BURGE
 ID LIP_BURGE STANDARD; PRT; 364 AA.
 AC P22088;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).
 GN LIPA.
 OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia.
 OX NCBI_TaxId=292;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 45-66.
 RC STRAIN=DSM 3959;
 RX MEDLINE=91100343; PubMed=1987151;
 RA Oerengen S., Skov K.W., Diderichsen B.,
 RT "Cloning, sequence, and expression of a lipase gene from Pseudomonas
 RT cepacia: lipase production in heterologous hosts requires two
 RT Pseudomonas genes.";
 RL J. Bacteriol. 173:559-567(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE=97184684; PubMed=9032073;
 RA Kim K.K., Song H.K., Shin D.H., Hwang K.Y., Suh S.W.,
 RT "The crystal structure of a triacylglycerol lipase from Pseudomonas
 RT cepacia reveals a highly open conformation in the absence of a bound
 RT inhibitor".
 RL Structure 5:173-185(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=97184685; PubMed=9032074;
 RA Schrag J.D., Li Y., Cygler M., Lang D., Burdorf T., Hecht H.-J.,
 RA Schmid R., Schomburg D., Rydel T.J., Oliver J.D., Strickland L.C.,
 RA Dunaway C.M., Larson S.B., Day J., McPherson A.,
 RT "The open conformation of a Pseudomonas lipase.";
 RL Structure 5:187-202(1997).
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
 CC fatty acid anion.
 CC -1- COPACITOR: Regulates calcium.
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS
 CC LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M58494; AAB50466.1; -.

```

DR PDB; 10L1; 15-MAY-97.
DR PDB; 2LIP; 12-MAR-97.
DR PDB; 3LIP; 16-JUN-97.
DR PDB; 1HOD; 22-AUG-01.
DR PDB; 4LIP; 19-AUG-98.
DR InterPro; IPRO00073; A/b hydrolase.
DR InterPro; IPRO08262; Lipase AS.
DR InterPro; IPRO00379; Ser_eester.
DR Pfam; PF00561; abhydrolase; 1.
DR PROSITE; PS00120; LIPASE_SER; 1.
DR Hydrolase; Lipid degradation; Signal; Calcium; 3D-structure.
FT SIGNAL 1 44
FT CHAIN 45 364 LIPASE. RELAY SYSTEM.
FT ACT_SITE 131 131 CHARGE RELAY SYSTEM.
FT ACT_SITE 286 286 CHARGE RELAY SYSTEM.
FT ACT_SITE 330 330 CHARGE RELAY SYSTEM.
FT TURN 48 49
FT STRAND 55 58
FT TURN 61 62
FT STRAND 66 67
FT TURN 68 70
FT STRAND 71 72
FT TURN 75 76
FT HELIX 77 83
FT TURN 84 85
FT STRAND 88 90
FT TURN 102 103
FT HELIX 105 120
FT TURN 121 121
FT STRAND 125 130
FT TURN 131 132
FT HELIX 133 143
FT HELIX 145 147
FT STRAND 148 154
FT TURN 158 159
FT HELIX 162 173
FT TURN 172 173
FT HELIX 175 196
FT TURN 178 194
FT HELIX 196 204
FT HELIX 204 210
FT TURN 211 211
FT HELIX 213 222
FT TURN 226 227
FT STRAND 228 228
FT TURN 231 232
FT STRAND 236 236
FT STRAND 240 242
FT STRAND 247 255
FT STRAND 258 264
FT TURN 265 266
FT STRAND 267 272
FT TURN 273 274
FT TURN 279 281
FT TURN 281 281
FT HELIX 282 285
FT HELIX 287 300
FT TURN 301 303
FT STRAND 307 307
FT STRAND 312 312
FT HELIX 313 316
FT STRAND 320 326
FT TURN 331 331
FT HELIX 332 334
FT TURN 335 339
FT TURN 343 344
FT HELIX 348 362
FT TURN 363 364
SQ SEQUENCE 364 AA; 37494 MW; E9CD2DBF555658E9 CRC64;

Query Match 9.1%; Score 85.5; DB 1; Length 364;
Best Local Similarity 27.4%; Pred. No. 3.7;
Matches 34; Conservative 20; Mismatches 43; Indels 27; Gaps 6;

```

```

QY 3 IIVHGL-----YMGIVMHPLSHRHKGYRQTITSYNSIAD-----EI 45
DB 55 IIVHGLSGTDKXAGVLEWYWG----IOEDIQQNGATVYVNLSPGSDDEPNRGEEL 109
QY 46 FRRLDRL--THASPNAIVGHSIGIVIRYLESAPSCETLSHVATIGSPLOGASIVNK 103
DB 110 LAVYKTVLATATGATKNLVGHSGGGL-SSRYVAAYVP--DLVASYTTITPRGSEPADF 166
QY 104 IEQL 107
DB 167 VQDV 170

RESULT 8
ATU1_YEAST STANDARD; PRT; 1216 AA.
ID ATU1_YEAST
AC P38360;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable copper-transporting ATPase (EC 3.6.3.4) (Cu(2+)-ATPase).
GN PCAL OR YBR295W OR YBR2112.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8288C;
RA MEDLINE=95274324; PubMed=7754711;
RX Rad M.R., Kirschner L., Hollenberg C.;
RT "A putative P-type Cu(2+)-transporting ATPase gene on chromosome II
of Saccharomyces cerevisiae."
RT Yeast 10:1217-1225(1994).
RL
CC -!- FUNCTION: Probably involved in copper transport and in the
CC regulation of cellular copper level.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Cu(2+)(In) = ADP + phosphate +
CC Cu(2+)(Out).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IB.
CC -!- SIMILARITY: Contains 1 HMA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z29332; CA82529.1; -.
CC EMBL; Z36164; CA85260.1; -.
CC PIR; S46177; S46177.
CC Germonline; 138838; -.
CC SCD; S0000499; PCAL.
CC GO; GO:0016020; C:membrane; IMP.
CC GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. .; IMP.
CC InterPro; IPRO06403; ATPase-IB1_Cu.
CC InterPro; IPRO06416; ATPase-IB_NY.
CC InterPro; IPRO01757; ATPase-IB-B2.
CC InterPro; IPRO01756; Cu ATPase.
CC InterPro; IPRO08250; E1-E2 ATPase reg.
CC InterPro; IPRO06121; HeavyMe transpct.
CC InterPro; IPRO05834; HeavyMe transpct.
CC InterPro; IPRO06191; Metal_bind.
CC Pfam; PF00403; HMA; 1.
CC Pfam; PF00702; Hydrolase; 1.
CC PRINTS; PR00119; CATATPASE.
CC PRINTS; PR00943; CUATPASE.
CC TIGRPFMS; TIGR01511; ATPase-IB1_Cu; 1.

```

DR TIGR01525; ATPase-1B hvy; 1.
 DR TIGR01494; ATPase P-Type; 2.
 DR PROSITE; PS00154; ATPase_E1_E2; 1.
 DR PROSITE; PS01047; HMA_1; PALISE_NEG.
 DR PROSITE; PS00846; HMA_2; 1.
 DR Hydrolyase; Copper transport; Transmembrane; Phosphorylation;
 KM Magnesium; ATP-binding; Metal-binding; Copper.
 FT DOMAIN 1 556
 FT TRANSMEM 557 578
 FT DOMAIN 579 592
 FT TRANSMEM 593 612
 FT DOMAIN 613 620
 FT TRANSMEM 621 641
 FT DOMAIN 642 659
 FT TRANSMEM 660 680
 FT DOMAIN 681 808
 FT TRANSMEM 809 831
 FT DOMAIN 832 847
 FT TRANSMEM 848 865
 FT DOMAIN 866 1161
 FT TRANSMEM 1162 1181
 FT DOMAIN 1182 1190
 FT TRANSMEM 1191 1209
 FT DOMAIN 1210 1216
 FT MOD RES 701 701
 FT METL 421 421
 FT METL 424 424
 FT METL 1107 1107
 FT METL 1111 1111
 SQ SEQUENCE 1216 AA; 131838 MW; B0BA4D60D75F9EA CRC64;
 Query Match
 Best Local Similarity 25.1%; Score 85; DB 1; Length 1216;
 Matches 43; Conservative 27; Mismatches 73; Indels 26; Gaps 8;
 QY 15 VNHPLSHRLKGYRTQTSYNSLAIDBAIFRLDRSLTASPNALVHSLG-----GL 69
 DB 893 VAHNTHSHVFD--KTGTLBGLTVVHEVTV--RQDR--HNSGSLGLTEGKHPVSM 944
 QY 70 VKRYLESRAPSCETLSHVVAI-GSPLOGASTVNIQLGVALGNSAFLGKENDP-- 126
 DB 945 AIASYLKEKGVSAQVNSKATGKRGVETS-----YSLKLGQSGNCRMTGHNDPVR 998
 QY 127 ---ESRYPOKSGSIAGTIPGLRSLILRDPDSDGTVEETKTAGMTDHI 174
 DB 999 KALGQGVSVFCFSVNGSV---TAVVALBDSILADAVSTINILRQGISLHI 1046
 RESULT 9
 P17B_ORYSA STANDARD; PRT; 268 AA.
 AC 04360;
 DT 01-NOV-1997 (Rel. 35; Created)
 DT 01-NOV-1997 (Rel. 35; Last sequence update)
 DT 10-OCT-2003 (Rel. 42; Last annotation update)
 DE P17B protein.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Indica-IR36;
 RA Reihmann C., Mauch F., Dudler R., Hofmann C.;
 RT "Characterization of a rice gene induced by Pseudomonas syringae pv.
 RT phyllis: requirement for the bacterial lema gene function.";
 RL Physiol. Mol. Plant Pathol. 46:71-81(1995).
 CC -1- INDUCTION: By infection of leaves with certain strains of P.
 CC syringae pv syringae.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL-ESTERASE/LIPASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; Z34271; CA84026.1; -
 CC EMBL; Z34270; CA84024.1; -
 CC PIR; S47087; S47087.
 CC HSSP; P52704; 1Q74.
 CC Gramene; Q43360; -
 CC InterPro; IPR000073; A/b_hydrolyase.
 CC InterPro; IPR000379; Ser_estrs.
 CC Pfam; PF00561; abhydrolyase; 1.
 CC Hydrolyase; Serine esterase.
 CC ACT_SITE 86 86
 CC ACT_SITE 218 218
 CC ACT_SITE 246 246
 CC ACT_SITE 246 246
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 SQ SEQUENCE 268 AA; 28635 MW; 9F4E67E1216560A1 CRC64;
 Query Match
 Best Local Similarity 25.3%; Score 80.5; DB 1; Length 268;
 Matches 47; Conservative 20; Mismatches 86; Indels 33; Gaps 6;
 QY 4 ILVHGVMGVLVNHPLSHRLKGYRTQTSY-----NSLAIDBAIFRLDRSLTHASF 58
 DB 12 ILVHGGLCHGKCMRYVVAALRAAGHATYALDMAASGAPARVDVTFEYSPLDDAVA 71
 QY 59 NA-----LVHSGLVGVIRKYLESRAPSCETLSHVVAIGSPLOGASTVNIQL-- 107
 DB 72 AAAPGRVLTVGHSHGSLVALAME-RPPDKVAALVFAAAMPVCGKMGVPTSEFMR 130
 QY 108 ----GL-----GVALGNSAFLGKENDDSRYQKSGSIAGTIPGLRSLILR----- 151
 DB 131 TAPEGLMDCEMVAINNQSQSGVALNIGPFLAQTQYQSPADLDALAKYLVAPNGQFMD 190
 QY 152 DPLDSD 157
 DB 191 DPLWMD 196
 RESULT 10
 PMPC_CHLMU STANDARD; PRT; 1460 AA.
 AC 09PVT1;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Probable outer membrane protein pmc precursor (Polymorphic membrane
 DE protein C).
 GN PMPC OR TC0695.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mopn / N199;
 RX MEDLINE=2015025; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
 RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- SUBCELLULAR LOCATION: Cell wall surface (elementary bodies)
 CC (potential).
 CC -1- SIMILARITY: Belongs to the PMP outer membrane protein family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: A5002338; AAF39511.1; -
DR PIR: D81675; D81675.
DR TIGR: TC0695; -
DR InterPro: IPR006315; Autotransport.
DR InterPro: IPR003368; Chlamydia_PMP.
DR Pfam: PF02415; Chlamydia_PMP; 2.
DR TIGRPFAM: TIGR01414; autotrans_bar1; 1.
DR TIGRPFAM: TIGR01376; POMP repeat; 8.
KM Outer membrane signal, Multigene family, Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1460 PROBABLE OUTER MEMBRANE PROTEIN PMPC.
SQ SEQUENCE 1460 AA; 154781 MW; C1F033E7907AC3AC CRC64;

Query Match 8.6%; Score 80.5; DB 1; Length 1460;
Best Local Similarity 26.1%; Pred. No. 48;
Matches 47; Conservative 25; Mismatches 69; Indels 39; Gaps 9;

QY 30 TQTSYNSLAIDDEAIFRRLDRSLTHAS---PNAIVGHSIGLVIKKYLESPAPSCETL 85
DB 490 TSPITQDASSHVAIFRSLAASSQSNSENI PNADGSTSAGDAGSSQSTPGSDSSI 549
QY 86 SHVVAIGSPLOGASVYKIEQL-GIGVALGNSAEFGKXEHDESRYPKSGSIAGTI--P 142
DB 550 NHYIG-GGAHYEAV--KIENLSGYGTFSSNNA---VDH-----QISSSTSDVIGGA 595
QY 143 LGRSLRLDPLDSDQTVVEET-----KIAG-----MTDHIKISTSYENA 184
DB 596 IYAKTSLTIDSGNSGCTIFSENTTSKSTTGQVAGGALFERSVITITPVPFSSKSAINA 655

RESULT 11
ASGC_SULTO
ID _ARGC_SULTO STANDARD; PRT; 349 AA.
AC 097655;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-
DB acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).
GN ARGC OR ST0195.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankat A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamaya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermocacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+) +
CC phosphate = N-acetyl-5-glutamyl phosphate + NADPH.
CC -1- PATHWAY: Arginine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily
CC 1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: APO0981; BAB65152.1; ALT_INIT.
DR HAMAP: MF_00150; -1
DR InterPro: IPR00706; AGPR_act_site.
DR InterPro: IPR000534; Semialdh_dh.
DR Pfam: PF01118; Semialdehyde_dh; 1.
DR Pfam: PF02774; Semialdehyde_dhc; 1.
DR ProDom: PD003765; AGPR_act_site; 1.
DR PROSITE: PS01224; AGCT, PALSE NEG.
KM Arginine biosynthesis; Oxidoreductase, NADP, Complete proteome.
FT ACT_SITE 150 150 BY SIMILARITY.
SQ SEQUENCE 349 AA; 38699 MW; EF1EE06BBA5847B CRC64;

Query Match 8.5%; Score 79.5; DB 1; Length 349;
Best Local Similarity 24.8%; Pred. No. 12;
Matches 38; Conservative 29; Mismatches 53; Indels 33; Gaps 9;

QY 19 LSHRLHKLGRQTGISYNSLAIDDEAIFRRLDRSL-----THASPNALVGHSLGGLYI 71
DB 76 LPHKV-SLEVPKILEMGIQVIDLSADFRUKPTLYKIYGYEHPPYD-LTKKAYGLPE 133
QY 72 KRYLESR-----APSCETLSHVVAIGSPLOGASVYKIEQLGIGVALGNSAEFGKXEH 125
DB 134 LHYELKNAKLISPGCNATATTLA-GAPLVASLLETKLIS-DYKVG-SSSGAKKPH 190
QY 126 DESRYPKSGSINGTIFLGRSLRLDPLDSDG 158
DB 191 G-SHHPERQVAIR-----PYEADG 208

RESULT 12
OM40_HUMAN
ID OM40_HUMAN STANDARD; PRT; 361 AA.
AC 096008; Q86VW4; Q8WY09; Q8WY10; Q8WY11; Q9BR95;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable mitochondrial import receptor subunit TOM40 homolog
DE (translocase of outer membrane 40 kDa subunit homolog) (Haymaker
DB protein) (p38.5).
GN TOMM40 OR TOM40 OR PEREC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=99449047; PubMed=10520737;
RA Freitas E.M., Zhang W.J., Lalonde J.P., Tay G.K., Gaudier S.,
RA Ashworth L.K., Van Bockmeer F.M., Dawkins R.L.;
RT "Sequencing of 42kb of the APO E-C2 gene cluster reveals a new gene:
RT PEREC1.";
RL DNA Seq. 9:89-101(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Yoshida K., Murray J.C.;
RT "A transcriptional map in the region of 19q13 derived using direct
RT sequencing and exon trapping.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lymphocytes;
RX MEDLINE=21610875; PubMed=11745481;
RA Das B., Tao S.-Z., Muhsitsky R., Norin A.J.;
RT "Genetic identity and differential expression of p38.5 (Haymaker) in
RT human malignant and non-malignant cells.";
RL Int. J. Cancer 94:800-806(2001).

[4] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RA TISSUE=Eye, lung, skin, testis, and uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stopleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Adnan R.D., Mullaly S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson W.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences".
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RL -1- FUNCTION: Essential for the import of protein precursors into the
 CC mitochondria (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC outer membrane (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O96008-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O96008-2; Sequence=VSP_008589, VSP_008590;
 CC Note=No experimental confirmation available;
 CC -1- SIMILARITY: Belongs to the Tom40 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

FT CONFLICT 49 49 S -> R (IN REF. 4; AAH06413).
 FT CONFLICT 219 219 A -> S (IN REF. 3; AAH46625).
 FT CONFLICT 220 220 A -> R (IN REF. 3; AAH46624).
 FT CONFLICT 258 258 N -> S (IN REF. 3; AAH46626).
 FT CONFLICT 297 297 T -> A (IN REF. 3; AAH46624).
 FT CONFLICT 313 313 L -> F (IN REF. 3; AAH46624).
 SQ SEQUENCE 361 AA; 37893 MW; CFE55E01F800D32 CRC64;
 Query Match 8.5%; Score 79.5; DB 1; Length 361;
 Best Local Similarity 23.2%; Pred. No. 12;
 Matches 43; Conservative 32; Mismatches 65; Indels 45; Gaps 8;
 QY 11 MGLVWEPYRSLH-KAGRTQTIISYNLAIDPAFLRDLSTPHASPNALVCHSLGL 69
 DB LAAQVTHQGPRLRRKMAIQTSKFVWQDGEIRGSDFTFAVTLGNPDVVG---SGI 216
 QY 70 VKRYIES-----RAPSCE-----TLSH---VVAIGSPQASIVN 102
 DB 217 LVAHYLGSTTPCLALGSLVYHRRPGEQTVWSLAKYTLNNWTLATVTLGQGMHATVYH 276
 QY 103 KT-EOLGVALGNABRGLEKHDESRYPQSGSIAGTIPGLRSLLRDPLSDGY- 160
 DB 277 KASDQNGV-----EFASITMOTDSVFGYQLDLPKANLLFKGSVDSNWIVG 325
 QY 161 -TVBE 164
 DB 326 ATLEK 330
 RESULT 13
 ACVS EMENT
 ID ACVS EMENT STANDARD; PRT; 3770 AA.
 AC P27742;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine synthase
 DE (EC 6.3.2.26) (Delta-(L-alpha-aminoacyl)-L-cysteiny-D-valine
 GN ACVA.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiaceae; Trichocomaceae; Emericella.
 OX NCBI_TaxId=162425;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=G191;
 RX MEDLINE=91286299; PubMed=2061333;
 RA Maccabe A.P., van Liemp H., Pallissa H., Unkles S.E., Riach M.B.R.,
 RA Pfeiffer E., von Doehren H., Kinghorn J.R.,
 RA Aspergillus nidulans. Molecular characterization of the acva gene
 RT encoding the first enzyme of the penicillin biosynthetic pathway".
 RL J. Biol. Chem. 266:12646-12654(1991).
 CC -1- FUNCTION: Each of the constituent amino acids of the tripeptide
 CC acv are activated as aminoacyl-adenylates with phosphate bonds
 CC formed through the participation of amino acid thioester
 CC intermediates.
 CC -1- CATALYTIC ACTIVITY: L-2-aminohexanoate + L-cysteine + L-valine
 CC + 3 ATP = N-(L-5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine +
 CC 3 AMP + 3 diphosphate.
 CC -1- COFACTOR: Contains 3 covalently bound phosphoguanethamines
 CC (Potential).
 CC -1- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
 CC -1- PTM: The N-terminus is blocked.
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X54853; CAA38631.1; -
 DR PIR; A40889; A40889.
 DR HSP; P14687; 1AMU.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR006163; Pp_bind.
 DR InterPro; IPR006162; Ppantn_S.
 DR InterPro; IPR000379; Ser_estr.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00501; AMP-binding; 3.
 DR Pfam; PF00668; Condensation; 3.
 DR Pfam; PF00550; pp-binding; 3.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
 DR PROSITE; PS00455; AMP_BINDING; 3.
 DR PROSITE; PS00075; ACP_DOMAIN; 3.
 KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
 Repeat; Phosphopantetheine.
 KW Repeat; 321 910
 FT REPEAT 1413 1993 DOMAIN 1 (ADIPATE-ACTIVATING).
 FT REPEAT 2494 3078 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 850 919 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
 FT DOMAIN 3087 3087 ACYL CARRIER (ACP) 3.
 FT BINDING 882 882 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 1965 1965 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 3050 3050 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 3770 AA; 422448 MW; C66B6D232A58CB0 CR664;

Query Match 8.5%; Score 79.5; DB 1; Length 3770;
 Best Local Similarity 25.3%; Pred. No. 1.7e+02;
 Matches 42; Conservative 31; Mismatches 62; Indels 31; Gaps 8;

QY 19 LSHRLKLGKGYRTQTSYNSLAIDDAIFR-RIDRSLTASPNALVGHSLGLVIRKYLE 77
 DB 1971 LSFKHEVGRITISV-----ALFRHRTISLAL-----LIMNVGD-----IQE 2010

QY 78 RAPSETTSHTVVAISPLQASIVNKKIOLGVALGNSAEPGLKEHDSERYPOKSGSI 137
 DB 2011 ITPVDYDNRKTAIVSPAQRLLFIHELH--GGGNAVYNIIDAFELPPYIDQSRVEALYTI 2068

QY 138 ACTIFLGRLRLPDLSDGT-----VTVEETKLAGMTDIAIST 178
 DB 2069 LSRHE-ALRTFLRD--QATGTFYOKITITDEAKCMILIKSNVST 2111

RESULT 14
 YCSE_BACSU STANDARD; PRT; 249 AA.
 AC P42962;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 10-OCT-2003 (rel. 42, Last annotation update)
 DE Hypothetical protein YCSE.
 OS YCSE OR BSU04040.
 GN Bacillus subtilis.
 OS Bacillus subtilis.
 NCBI_Taxid:1423;
 OX NCBI_Taxid:1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Akaagawa E., Kurita K., Sugawara T., Nakamura K., Yamane K.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN:168;
 RM MEDLINE=97124189; PubMed=8969502;

RA Yamane K., Kumano M., Kurita K.;
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis
 RT chromosome: determination of the sequence of a 146 kb segment and
 RT identification of 113 genes.";
 RL Microbiology 142:3047-3056 (1996).
 RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN:168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Boudreau S., Boudreau L., Brans A., Braun M., Brignall S.C., Bron S.,
 RA Brouillet S., Bruchet C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.D., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertan K.D., Errington J., Fabre C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gim S.Y., Glaeser P., Goffeau A., Goughly E.C., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech U., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaer-Bianhard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Meliadi R.P., Mizuno M., Noesti D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawa K., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,
 RA Sekiguchi J., Sekowka A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Toseco V., Uchiyama S., Vandepol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Welzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256 (1997).
 CC -1- SIMILARITY: BEYONDS TO THE COF/YBHA/YIDA/YTGL (E.COLI) / YCSE/YXEH
 CC (B.SUBTILIS) FAMILY.
 CC CC
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

Query Match 8.4%; Score 79; DB 1; Length 249;
 Best Local Similarity 26.4%; Pred. No. 8.7;
 Matches 24; Conservative 22; Mismatches 27; Indels 18; Gaps 4;

QY 43 EAIFRRLRLSLTHASPNALVGHSLGVLTKRYLSRAPSC-----TLSHVVALGSP 95
 DB 150 BELRTKSSSEITNSPTNIEVNALG-----INKAAPCGYKGLGTWENVAMWMDSL 202
 QY 96 QGASIVNKIEQLGVALGNSAERGLKEHDD 126
 DB 203 NDIMW---IKKAGLGVMAGNAGDI-VKRTAD 229

RESULT 15
 1963 MYCTU STANDARD; PRT; 266 AA.
 AC P71547;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein RV0963c/MT0992/MB0988C.
 GN RV0963C OR MT0992 OR MTCY10D7.11 OR MB0988C.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_taxid=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmeier K., Gas S., Barry C.B. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean L.A., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.B., Taylor K., Whitehead S., Barrell B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwin M., Haft D., Hickey E.,
 Kojanay J.F., Nelson W.C., Mayam L.A., Ermolaeva M., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12786972;
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
 Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -I- SIMILARITY: SOME. TO M.TUBERCULOSIS RV2079.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; Z79700; CAB02008.1; -
 DR EMBL; AE006963; AKA5240.1; -
 DR EMBL; EX248337; CAD93849.1; -

DR PIR; A70718; A70718.
 DR TIGR; MT0992; -
 DR Tuberculet; RV0963c; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 266 AA; 28069 MW; E1FA43081A6B61AD CRC64;
 Query Match 8.4%; Score 79; DB 1; Length 266;
 Best Local Similarity 26.2%; Pred.No. 9.3;
 Matches 42; Conservative 19; Mismatches 51; Indels 48; Gaps 9;
 QY 26 LGYRT---QTISYNSLAIDDEAIFRRLDRSL---THASPNALV---GSLGGLVTKRYLE 76
 DB 122 LGYDAPDGLKDVVHMSARDAGAPLNRFKGLAATTNVSDDCHITAFGHSYGSGLVLSLQ 181
 QY 77 SRAPSCETLSHYVALGSPVQASIVNKIEQLGVALGNS-AERGLKEHDDSRYPQKSG 135
 DB 182 QGAP---VSDVVLVGGSP--GTGLTHASQ---LGVEPGHAFYMGVNDH----- 221
 QY 136 SIAGTIP-----LGLRSLRLRDLPLDSPG 158
 DB 222 -VANTIPERGAFGSAPQDVPGMTQLSVNTGLAPGLDGG 260

Search completed: April 29, 2004, 11:36:01
 Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:24:52 ; Search time 41 Seconds
(without alignments)
1423.680 Million cell updates/sec

Title: US-10-603-260-5
Perfect score: 936
Sequence: 1 MGIILVGLVWGLVWHP...KIAQMDHIAISTSYENAV 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_rhiz:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 577.5 | 61.7 | 215 | 16 | Q8D5U8 |
| 2 | 566.5 | 60.5 | 209 | 16 | Q87GR6 |
| 3 | 532 | 56.8 | 205 | 16 | Q9KN73 |
| 4 | 179.5 | 19.2 | 199 | 16 | Q8PEE2 |
| 5 | 170.5 | 18.2 | 224 | 16 | Q89S87 |
| 6 | 155 | 16.6 | 98 | 2 | Q87026 |
| 7 | 136.5 | 14.6 | 328 | 16 | Q7UNU3 |
| 8 | 136 | 14.5 | 180 | 16 | Q8PHU7 |
| 9 | 121.5 | 13.0 | 249 | 16 | Q55748 |
| 10 | 112.5 | 12.0 | 251 | 16 | Q9PHH3 |
| 11 | 111.5 | 11.9 | 254 | 2 | Q8KU33 |
| 12 | 107 | 11.4 | 256 | 16 | Q98KTS |
| 13 | 105.5 | 11.3 | 233 | 16 | Q98UB5 |
| 14 | 105.5 | 11.3 | 288 | 16 | Q8UXB1 |
| 15 | 104.5 | 11.2 | 243 | 16 | Q92XY1 |
| 16 | 103.5 | 11.1 | 202 | 16 | P73372 |

| | | | | | |
|----|-------|------|-----|----|--------|
| 17 | 102.5 | 11.0 | 256 | 16 | Q89ML2 |
| 18 | 99.5 | 10.6 | 211 | 16 | Q8YK63 |
| 19 | 97 | 10.4 | 293 | 2 | Q9EV86 |
| 20 | 97 | 10.4 | 309 | 16 | Q87QG8 |
| 21 | 97 | 10.4 | 623 | 16 | Q8F4V3 |
| 22 | 96.5 | 10.3 | 309 | 2 | P72172 |
| 23 | 96.5 | 10.3 | 309 | 16 | Q9HJZ7 |
| 24 | 96 | 10.3 | 327 | 16 | Q9ZBN4 |
| 25 | 95.5 | 10.2 | 340 | 16 | Q8X873 |
| 26 | 95.5 | 10.2 | 340 | 16 | Q8FCX8 |
| 27 | 95.5 | 10.2 | 340 | 16 | Q83JB6 |
| 28 | 95 | 10.1 | 255 | 16 | Q8UKN4 |
| 29 | 95 | 10.1 | 262 | 16 | Q7UOF1 |
| 30 | 94.5 | 10.1 | 259 | 16 | Q8XWZ0 |
| 31 | 94.5 | 10.1 | 261 | 16 | Q83XN6 |
| 32 | 94.5 | 10.1 | 338 | 2 | Q7X568 |
| 33 | 94.5 | 10.1 | 340 | 16 | Q7UAR6 |
| 34 | 94 | 10.0 | 332 | 16 | Q916C6 |
| 35 | 93.5 | 10.0 | 336 | 13 | Q7ZTW6 |
| 36 | 93 | 9.9 | 302 | 16 | Q8DA61 |
| 37 | 92.5 | 9.9 | 258 | 16 | Q81EC2 |
| 38 | 92.5 | 9.9 | 284 | 16 | Q8KQJ2 |
| 39 | 92 | 9.8 | 290 | 16 | Q8S295 |
| 40 | 91.5 | 9.8 | 344 | 13 | Q7E314 |
| 41 | 91 | 9.7 | 242 | 10 | Q9FF27 |
| 42 | 91 | 9.7 | 266 | 16 | Q89R93 |
| 43 | 90.5 | 9.7 | 266 | 16 | Q82HP7 |
| 44 | 90.5 | 9.7 | 300 | 16 | Q92D71 |
| 45 | 90 | 9.6 | 195 | 16 | Q7U656 |

ALIGNMENTS

RESULT 1

| ID | Q8D5U8 | PRELIMINARY; | PRT; | 215 AA. |
|----|--|--------------|------|---------|
| AC | Q8D5U8; | | | |
| DT | 01-MAR-2003 (TREMBLrel. 23, Created) | | | |
| DT | 01-MAR-2003 (TREMBLrel. 23, Last sequence update) | | | |
| DT | 01-OCT-2003 (TREMBLrel. 25, Last annotation update) | | | |
| DE | Predicted acetyltransferase/hydrolase. | | | |
| GN | V20805. | | | |
| OS | Vibrio vulnificus. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; | | | |
| OC | Vibrionaceae; Vibrio. | | | |
| OX | NCBI_TaxID=672; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=CMCP6; | | | |
| RA | Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H., Choy H.E.; | | | |
| RT | "Complete genome sequence of Vibrio vulnificus CMCP6."; | | | |
| RL | Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases. | | | |
| DR | EMBL: A016810; A007231.1; | | | |
| DR | GO: GO:0016787; F:hydrolase activity; IEA. | | | |
| DR | GO: GO:0016740; F:transferase activity; IEA. | | | |
| DR | InterPro: IPR000379; Ser. esters. | | | |
| KW | Hydrolase; Transferase; Complete proteome. | | | |
| SQ | SEQUENCE 215 AA; 23875 MW; 5AD7E9FE1B690E2E CRC64; | | | |

Query Match 61.7%; Score 577.5; DB 16; Length 215;
Best Local Similarity 64.6%; Pred. No. 1.2e-45;
Matches 113; Conservative 28; Mismatches 33; Indels 1; Gaps 1;

| | | |
|----|---|-----|
| QY | 1 MGIILVGLVWGLVWHP...KIAQMDHIAISTSYENAV 185 | 60 |
| DB | 6 MGIILVGLVWGLVWHP...KIAQMDHIAISTSYENAV 185 | 65 |
| QY | 61 LVGSHSLGIMIKHYLASRP...QIISHVAVIGSPFLKASIVTKIOELGAMLVNSPKHG 125 | 120 |
| DB | 66 LVGSHSLGIMIKHYLASRP...QIISHVAVIGSPFLKASIVTKIOELGAMLVNSPKHG 125 | 125 |


```

QY 3 IIVHGGIYMGGLWHPFSHLLHKGRTCTISNLSAIDEAFRDLSDSLTHASPNLV 62
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1 MLTVHGIWNTAWMLPLFARLRFTAGMTPLFGVAFRGSGPAQVPLRLRLATPAFLV 60

QY 63 GHSIGGIYKRYLESRAFCETLSHYVAAGSPQASINYNKEIOLGVALNSAFGLX 12
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 61 CHSIGGMALQNG-QAPFL-PTQRYVCCSPICSGAARGLARGGHMAKRSALILR 11

QY 123 --EHDSSRRPQSGSAGTIPGLBSLLRPDLDDGYVYBEYKIAMTDHIAISTT 179
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 119 GFTIMDA---AQGGQAGCVPRGRLGRLWAPLQAGSDGVYAGDEGLPELHIDHCVQAS 174

```

| | |
|----------|--------------|
| RESULT 5 | |
| Q89SS7 | |
| ID | Q89SS7 |
| | PRELIMINARY; |
| | PRT; 224 AA |

DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE B112323 protein.
 GN B112323.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_Taxid=575;
 RX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=USDA 110;
 RC MEDLINE=22484998; PubMed=12597275;
 RX Xanezo T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Watanabe A., Ideesawa K., Iriuchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada
 RA Tabata S.;
 RT "complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RT DNA Res. 9:189-197(2002).
 RL EMBL; AP005943; BAC47588.1; -;
 DR GO; GO:0003824; Fcatalytic activity, IEA.
 DR InterPro; IPR000379; Set_restrict.
 KW Complete proteome.
 SQ SEQUENCE 224 AA; 23746 MW; D58F7229BED11465 CRC64;

| | | | | |
|--------------------------|--------|--------------------|------------|-------------|
| Query Match | 18.2%; | Score 170.5; | DB 16; | Length 224; |
| Best Local Similarity | 29.1%; | Pred. No. 9.5e-08; | | |
| Matches 57; Conservative | 34; | Mismatches 74; | Indels 31; | Gaps 8 |

```

QY      3 IIVVHGIVMGMVWHPLSHRKHLGRTQTIYSN-----LAIDSE---AIFRDLDSLT 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      15 VVLIHGIAKRSASLTTLERALRASGFTTLVNDVYSRSKPIALADHDIPALARAEND-- 72
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      55 HASPNMLVGHSLGLVYIKRYLESRAPECSCTLSHVAALGPSLOGASIVNKIEQL----- 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      73 --APLPHVHNSMGGLVARAYIAHRRA--RLARVMLGTPNSGSEVADLLGGLLPYAFY 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      108 -GLGVALGN---SAPFGLKEHDDSEKRYPOKSSISIGTIPLGRSLILRLDPDSIGTYVE 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      129 GPAGLEITETTPAPAPALPAID---YP--VGVIAGNRFIDPVAGLFVLPPENDGRSVQ 182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      164 ETKIAGTCHIAISTT 179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      183 STMLAGMTDHHVVVKAS 198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

| | | |
|--------|---------------|---|
| RESULT | 6 | |
| 087026 | | |
| AC | 087026 | PRELIMINARY; |
| ID | 087026 | PRT; 98 AA. |
| DT | 01-NOV-1998 | (T-EMBLrel. 08 Created) |
| DT | 01-NOV-1998 | (T-EMBLrel. 08, Last sequence update) |
| DT | 01-NOV-1998 | (T-EMBLrel. 08, Last annotation update) |
| DE | Z10r protein. | |
| NN | Z10r. | |

OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z17561;
RA Fallarino A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
SQ EMBL; AJ231080; CAA1312.1; -
SQ SEQUENCE 98 AA; 11040 MW; 4350024076808F9C CRC64;

| | | | | |
|-----------------------|-------|--------------|---------|---------------------------------|
| Query Match | 16.6% | Score 155 | DB 2 | length 98 |
| Best Local Similarity | 50.0% | Pred. No. | 8.5e-07 | |
| Matches | 32 | Conservative | 10 | Mismatches 22; Indels 0; Gaps 0 |

OY 113 LGSNAFFGKXEHDESRYPQKSGSIAGTIPGLSLRLPDLSDGVTVAEFRKMGTD 172
 | : : : | : : : | : : : | : : : |
Db 2 LGAAHLYGQLHQDSWELPQRUCIACILRLRRGFPIILLGGSGMCDGYTVTAERQISGMTD 61

OY 173 HIAI 176
 | : :
Db 62 HLTL 65

| RESULT 7 | |
|----------|--------------------------|
| ID | PRELIMINARY; PRT; 328 AA |
| Q7UNU3 | |
| Q7UNU3 | |
| Q7UNU3 | |

DT 01-0CT-2003 (TREMBlrel, 25, Created)
DT 01-0CT-2003 (TREMBlrel, 25, Last sequence update)
DT 01-0CT-2003 (TREMBlrel, 25, Last annotation update)
DE Hypothetical protein.
GN B8374.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22755913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleener H., Amani R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp
RT strain 1." Acad. Sci. U.S.A. 100:8298-8303(2003).
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
RW EMBL, BX294146, CAD75322.1, --
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 328 AA; 36089 MW; 6F831D70CFE9F5 CRC64;

| | | | | |
|-----------------------|-----------------|------------------|-----------|------------|
| Query Match | 14.6% | Score 136.5 | DB 16 | length 328 |
| Best Local Similarity | 27.8% | Pred. No. 0.0024 | | |
| Matches 52 | Conservative 34 | Mismatches 74 | Indels 27 | Gaps 9 |

```
QY      IIIVHVLVNHGLVMHPHSRLKHKGYRQTSTSY-----NSLAIDDAIPRRLLDRSRSTA 56
Db      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
        118 VMLTHGMEETDNCKMSLETKLHAESTYE-QTIRFGYSTRTSLSA-ESMAALRDYLENOSSD 175
QY      57 SPNALGHSIGGLVIKRYLESRAPECE-----TLSHVVAIGSPLOGASIVNKIEQLGL- 109
Db      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
        176 AEPAPFVGHSNGINIVRHILIGDLQADGDPKNLPLFKSMWMLGPNNQGAIALARSLAPGVAF 235
QY      110 GVALGNISA-EFGKEKHEDBSRY---PQKSGSIAGTIPLGLRSLLLRPPL---DSGGTYIV 162
Db      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
        236 GLVAGSGAMELGTRSNWSBESHLPFPFAIVAQRV-----APPANPLVDGGDFVSVL 290
QY      163 EETKING 169
Db      :|::|||
        291 EEAOLEG 297
```

RESULT 8
 O9PHU7 PRELIMINARY; PRT; 180 AA.
 AC O9PHU7;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE Hypothetical protein XAC3152.
 GN XAC3152.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=2022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitelio C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 RA Camarotte G., Camavari F., Cardozo J., Chamberg F., Ciapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.L., Curcio-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
 RA Fomghieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
 RA Setudal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RT Nature 417:459-463(2002).
 RL EMBL; AE011959; AAM37996.1;
 DR EMBL; AE011959; AAM37996.1;
 KM Hypothetical protein, Complete proteome.
 SQ SEQUENCE 180 AA; 19018 MM; 6DDB43872694A746 CRC64;
 Query Match 14.5%; Score 136; DB 16; Length 180;
 Best Local Similarity 24.9%; Pred. No. 0.00011;
 Matches 44; Conservative 31; Mismatches 74; Indels 28; Gaps 4;
 QY 3 ILVVGIVMGLVWMLPRLHKLGYRTOTISYNLSALDDAIPRLDRSLTHSPNAV 62
 DB 1 VLVVGIVMGLVWMLPRLHKLGYRTOTISYNLSALDDAIPRLDRSLTHSPNAV 60
 QY 63 GHSGLGLVTKRYLESRAPECELTSHVAIGSPLOQASIVNKIEQLGLVALGNSA---EF 119
 DB 61 GHSGLGLVTKRYLESRAPECELTSHVAIGSPLOQASIVNKIEQLGLVALGNSA---EF 118
 QY 120 GLKXHDSDSRYPQKSGS--IAGTTPGLRSLRLRPLDSDGVTVEETKAGTDTIAT 176
 DB 119 GPARWDGQALTIQVAGNVA-----LAETRLPLGLRDHCIV 152
 RESULT 9
 O55748 PRELIMINARY; PRT; 249 AA.
 AC O55748;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein slr0421.
 GN slr0421.
 OS Synecocystis sp. (Strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 OX NCBI_TaxID=1148;
 RN NCBI_TaxID=1148;
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC 6803;
 RA Tabata S.;
 RL Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Saruka T., Miyajima N.,
 RA Sugita M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 644 to 924 of the genome."
 RL DNA Res. 2:153-166(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hikosawa M., Sugita M., Sasaoka S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Ntaro K., Okumura S.,
 RA Shimizu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions."
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D64002; BAA10393.1;
 DR PIR; S76547; S76547.
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR InterPro; IPR000379; Ser_ester.
 KM Hypothetical protein, Complete proteome.
 SQ SEQUENCE 249 AA; 27534 MM; F36D75D573232008 CRC64;
 Query Match 13.0%; Score 121.5; DB 16; Length 249;
 Best Local Similarity 32.7%; Pred. No. 0.004;
 Matches 37; Conservative 20; Mismatches 39; Indels 11; Gaps 5;
 QY 60 ALVGHSLGLVTKRYLESRAPECELTSHVAIGSPLOQASIVNKIEQLGLVALGNSA 118
 DB 72 SVVGFPMGGLVWMLPRLHKLGYRTOTISYNLSALDDAIPRLDRSLTHSPNAV 128
 QY 119 GLKXHDSDSRYPQKSGS--IAGTTPGLRSLRLRPLDSDGVTVEETKAGTDTIAT 169
 DB 129 LGRRQYLAETKTKIPLVIVAGSVNGS-----GDGVIPRESIRLNG 170
 RESULT 10
 O9PHH3 PRELIMINARY; PRT; 251 AA.
 AC O9PHH3;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein Xfa0032.
 GN XFA0032.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RP SEQUENCE FROM N.A.
 RC STRAIN=9945C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.U.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarado R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britton M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carrer D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Fröhne M., Furlan L.R.,
 RA Gardner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes S.A., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

| RESULT 11 | | | |
|-----------|--|--------------|--------------|
| ID | Q8KU33 | PRELIMINARY; | PRT; 254 AA. |
| AC | Q8KU33; | | |
| DT | 01-OCT-2002 (TREMBLrel. 22, Created) | | |
| DT | 01-OCT-2002 (TREMBLrel. 22, Last sequence update) | | |
| DT | 01-OCT-2003 (TREMBLrel. 25, Last annotation update) | | |
| DE | Putative lipase LlpB. | | |
| OS | Legionella pneumophila. | | |
| OC | Bacterial Proteobacteria; Gammaproteobacteria; Legionellales; | | |
| OC | Legionellaceae; Legionella. | | |
| OX | NCBI_TaxId=446; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=130b. | | |
| FX | MEDLINE=22095963; PubMed=12101309. | | |
| RA | Aragón V., Rossier O., Cianciocto N.P., | | |
| RT | "Legionella pneumophila genes that encode lipase and phospholipase C | | |
| RT | activities."; | | |
| RL | Microbiology 148:2223-2231(2002). | | |
| DR | EMBL; AF54864; AAM73853.1; "- | | |
| DR | GO: GO:0003824; Catalytic activity; IEA. | | |
| DR | InterPro; IPR000379; Set_SecrB | | |
| QO | SEQUENCE 254 AA; 28850 MW; 999C2E079C102011 CRC64; | | |

| | | | | |
|-----------------------|------------------|------------------|------------|-------------|
| Query Match | 11.9%; | Score 111.5; | DB 2; | Length 254; |
| Best Local Similarity | 24.4%; | Pred. No. 0.035; | | |
| Matches 47; | Conservative 30; | Mismatches 81; | Indels 35; | Gaps 8; |

| RESULT 12 | | | |
|-----------|---|--------------|--------------|
| ID | Q98KT5 | PRELIMINARY; | PRT; 256 AA. |
| AC | Q98KT5; | | |
| DT | 01-OCT-2001 (TrEMBLrel. 18, Created) | | |
| DT | 01-OCT-2001 (TrEMBLrel. 18, Last sequence update) | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | |
| DE | Hypothetical protein ml1329. | | |
| GN | ML1329. | | |
| OS | Rhizobium loti (Mesorhizobium loti). | | |
| OC | Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; | | |
| OC | Phyllobacteriaceae; Mesorhizobium. | | |
| OX | NCBI_TaxID=361; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=MAFF30309; | | |
| RX | MEDLINE=11082930; PubMed=11214968; | | |
| RA | Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S., | | |
| RA | Watanabe A., Iiesawa K., Ishikawa A., Kawashima K., Kimura T., | | |
| RA | Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., | | |
| RA | Mochozuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M., | | |
| RA | Takeuchi C., Yamada M., Tabata S.; | | |
| RT | "Complete genome structure of the nitrogen-fixing symbiotic bacterium | | |
| RT | Mesorhizobium loti." | | |
| RL | DNA Res. 7:331-338(2000). | | |
| DR | EMBL; AP002997; BAB48729.1; . | | |
| DR | GO; GO:0003824; Catalytic activity; IEA. | | |
| DR | InterPro; IPR000379; Ser esters. | | |
| KW | Hypothetical protein; Complete proteome. | | |
| SQ | SEQUENCE 256 AA; 26643 MW; 7AC63AB6830E716C CRC64; | | |

| | | | | | |
|-----------|-------------------------------|---|-------------------------|--------------------------|--------------------------------------|
| | | | Query March | 11.4%; Score 107; DB 16; | length 256; |
| | | | Best Local Similarity | 24.6%; | Pred. No. 0.093; |
| | | | Matches | 42; Conservative | 28; Mismatches 67; Indels 34; Gaps 6 |
| Oy | 3 | IIVNHGLVHAGLVHPLESHRLHLKGIPTQTS--YNSLAIDDELFRRLDRSLTHASPN | A 60 | | |
| Db | 35 | IIVNHGAFVDQTSMKRFVADILTKKGYNVTILENPPTSIAADVDAATKALAK--QNGKT | V 91 | | |
| Oy | 61 | LVSHTSGVLVR-----YLERSAPSC-ETLSHVVAISPLQASIVNKRIBLG | G 108 | | |
| Db | 92 | LVSMSGVVVITIQAGDDPVSAIYYVSAFAPDVGBSIALSKSPATEGAALAIHPDEKN | E 151 | | |
| Oy | 109 | LGVALGNSEAFPELKEHDRESRYPCQSSTACTPIGLARILLRPPLDSDG | N 159 | | |
| Db | 152 | LTV-----DKEYFP---SAVAADLPETIMESLANHQPLINRT | T 185 | | |
| RESULT 13 | | | | | |
| ID | 09BJB5 | PRELIMINARY; | PRT; | 233 AA. | |
| AC | 09BJB5; | | | | |
| DT | 01-OCT-2001 | (T-EMBLrel. 18, | Created) | | |
| DT | 01-OCT-2001 | (T-EMBLrel. 18, | Last sequence update) | | |
| DT | 01-OCT-2003 | (T-EMBLrel. 25, | Last annotation update) | | |
| DE | Hypothetical protein ml12018. | | | | |

GN ML2018.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kato T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaomoto S.,
 Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AB002998; BAB49251.1;
 DR GO; GO:0003824; P: catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR000379; Ser esters.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 23925 MW; F93FE565C548D63 CRC64;

Query Match 11.3%; Score 105.5; DB 16; Length 233;
 Best Local Similarity 35.1%; Pred. No. 0.11;
 Matches 40; Conservative 20; Mismatches 31; Indels 23; Gaps 8;
 Db 3 IILVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 59
 9 VVLVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 63
 60 A-LVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 100
 64 VTLVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 115

RESULT 14
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kato T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaomoto S.,
 Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AB002998; BAB49251.1;
 DR GO; GO:0003824; P: catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR000379; Ser esters.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 23925 MW; F93FE565C548D63 CRC64;

Query Match 11.3%; Score 105.5; DB 16; Length 233;
 Best Local Similarity 35.1%; Pred. No. 0.11;
 Matches 40; Conservative 20; Mismatches 31; Indels 23; Gaps 8;
 Db 3 IILVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 59
 9 VVLVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 63
 60 A-LVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 100
 64 VTLVHGSLGVLKRLSHRLKLGRTQISYN--SLAIDDEAIFRRLDRSLTHASPN 115

RESULT 14
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kato T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaomoto S.,
 Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AB002998; BAB49251.1;
 DR GO; GO:0003824; P: catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR000379; Ser esters.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 23925 MW; F93FE565C548D63 CRC64;

RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Nollan C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
 RA Claeys C., Slater S.,
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328 (2001).
 DR EMBL; AB009944; AAL45901.1; ALT_INIT.
 DR EMBL; AB007891; AAK90583.1;
 DR FPI; AG3185; AG3185.
 DR GO; GO:0046821; C: extrachromosomal DNA; IEA.
 DR GO; GO:0003824; P: catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR000379; Ser esters.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KM Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 288 AA; 30807 MW; 0C586AA4578FEAB CRC64;

Query Match 11.3%; Score 105.5; DB 16; Length 288;
 Best Local Similarity 28.3%; Pred. No. 0.15;
 Matches 45; Conservative 19; Mismatches 68; Indels 27; Gaps 6;
 Db 3 IILVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 60
 63 VVLVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 119
 61 LVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 118
 120 LVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 168
 119 FGLKHHDESRYPQKSGSLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 157
 169 FVLETTDGPFGY-----VSPAKKSGFADVSDAD 198

RESULT 15
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kato T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaomoto S.,
 Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AB002998; BAB49251.1;
 DR GO; GO:0003824; P: catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR000379; Ser esters.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 243 AA; 25940 MW; 59A484BD1DBA850 CRC64;

Query Match 11.2%; Score 104.5; DB 16; Length 243;
 Best Local Similarity 28.9%; Pred. No. 0.15;
 Matches 45; Conservative 19; Mismatches 68; Indels 27; Gaps 6;
 Db 3 IILVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 60
 63 VVLVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 119
 61 LVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 118
 120 LVHGSLGVLKRLSHRLKLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 168
 119 FGLKHHDESRYPQKSGSLGRTQIS--VNSLAIDDEAIFRRLDRSLTHASPN 157
 169 FVLETTDGPFGY-----VSPAKKSGFADVSDAD 198

RESULT 15
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kato T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaomoto S.,
 Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AB002998; BAB49251.1;
 DR GO; GO:0003824; P: catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR000379; Ser esters.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 243 AA; 25940 MW; 59A484BD1DBA850 CRC64;

Matches 55; Conservative 19; Mismatches 77; Indels 39; Gaps 10;

```

QY      3 IILVHGLMHHGLVMMHPLSHRLHLKGYRQTSYNSLAI--DDEAIFRR-IDRSLTHASPN 59
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     18 VVLVHGAFADSGMKGVYDNLTKRGYV-TIVQNPFTSLIEDDVATRRALER--QDGPV 73
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY     60 ALVGHSLGLVTKRYLBSRAPSCETLSHVVAIGSPLOGASIVNKIEQLGLGVALGNSAEF 119
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     74 ILVGHSWGQTVITE--TGIDPKVAGLVVVAL-SPDAGETTAQOE-----GFAPAAEF 124
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    120 GLKEHDDSRYPQKSGSIAGTIFPLGLRSLLRDPIDSDGTIVTEETKIAGMTD----HI 174
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    125 VIETTKDGRGY-----VSPAKFKAGFAHDVSDAD-----VAFMRDAQVPIM 166
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    175 AISTSYENA 184
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    167 SAFATKLENA 176
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: April 29, 2004, 11:36:59
 Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:34:34 / Search time 24 Seconds
(without alignments)
397.950 Million cell updates/sec

Title: US-10-603-260-5
Perfect score: 936
Sequence: 1 MQLIVHGLVHGLVHGLHPLS.....KIAGMTDHAISTSYENAV 185

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.rep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.rep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.rep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.rep.*
5: /cgn2_6/ptodata/2/1aa/PTCUS.COMB.rep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.rep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1 | 97 | 10.4 | 277 | US-09-111-556A-3 | Sequence 3, Appl1 |
| 2 | 97 | 10.4 | 277 | US-08-360-758-3 | Sequence 3, Appl1 |
| 3 | 96.5 | 10.3 | 351 | US-09-252-991A-27337 | Sequence 27337, A |
| 4 | 96.5 | 10.3 | 453 | US-09-328-352-5069 | Sequence 5069, Ap |
| 5 | 96 | 10.3 | 316 | US-09-107-532A-4594 | Sequence 4594, Ap |
| 6 | 96 | 10.3 | 358 | US-08-034-650-10 | Sequence 10, Appl |
| 7 | 96 | 10.3 | 358 | US-08-449-015-10 | Sequence 10, Appl |
| 8 | 92.5 | 9.9 | 333 | US-09-543-681A-6874 | Sequence 6874, Ap |
| 9 | 91.5 | 9.8 | 262 | US-08-603-359A-38 | Sequence 38, Appl |
| 10 | 91 | 9.7 | 313 | US-08-733-412-2 | Sequence 2, Appl1 |
| 11 | 90.5 | 9.7 | 363 | US-08-978-589A-2 | Sequence 2, Appl1 |
| 12 | 90.5 | 9.7 | 363 | US-09-336-601-1 | Sequence 1, Appl1 |
| 13 | 90.5 | 9.7 | 363 | US-09-219-120-2 | Sequence 2, Appl1 |
| 14 | 89 | 9.5 | 344 | US-09-252-991A-27871 | Sequence 27871, A |
| 15 | 88.5 | 9.5 | 151 | US-08-637-759B-349 | Sequence 349, App |
| 16 | 88.5 | 9.5 | 151 | US-08-871-355A-349 | Sequence 349, App |
| 17 | 88.5 | 9.5 | 151 | US-09-201-945-349 | Sequence 349, App |
| 18 | 88.5 | 9.5 | 652 | US-09-134-001C-3517 | Sequence 3517, Ap |
| 19 | 85 | 9.1 | 303 | US-09-489-039A-10931 | Sequence 10931, A |
| 20 | 84.5 | 9.0 | 774 | US-09-252-991A-29669 | Sequence 29669, A |
| 21 | 82 | 8.8 | 244 | US-09-198-452A-360 | Sequence 360, App |
| 22 | 81 | 8.7 | 922 | US-09-141-206-6 | Sequence 6, Appl1 |
| 23 | 80.5 | 8.6 | 312 | US-09-134-000C-4460 | Sequence 4460, Ap |
| 24 | 80.5 | 8.6 | 720 | US-09-252-991A-19581 | Sequence 19581, A |
| 25 | 79 | 8.4 | 270 | US-09-071-035-188 | Sequence 188, App |
| 26 | 79 | 8.4 | 292 | US-09-071-035-186 | Sequence 186, App |
| 27 | 79 | 8.4 | 299 | US-09-584-568C-6 | Sequence 6, Appl1 |

| | | | | | |
|----|------|-----|-----|----------------------|--------------------|
| 28 | 79 | 8.4 | 527 | US-09-252-991A-32457 | Sequence 32457, A |
| 29 | 78.5 | 8.4 | 287 | US-09-584-568C-4 | Sequence 4, Appl1 |
| 30 | 78.5 | 8.4 | 294 | US-09-489-039A-13295 | Sequence 13295, A |
| 31 | 78.5 | 8.4 | 313 | US-09-252-991A-18084 | Sequence 18084, A |
| 32 | 78.5 | 8.4 | 511 | US-09-489-039A-7682 | Sequence 7682, Ap |
| 33 | 78.5 | 8.4 | 603 | US-09-976-594-260 | Sequence 260, App |
| 34 | 78.5 | 8.4 | 621 | US-09-121-964-8 | Sequence 8, Appl1 |
| 35 | 77.5 | 8.4 | 621 | US-09-155-078-2 | Sequence 2, Appl1 |
| 36 | 77.5 | 8.3 | 568 | US-09-252-991A-29237 | Sequence 29237, A |
| 37 | 77.5 | 8.3 | 638 | US-09-252-991A-17957 | Sequence 17957, A |
| 38 | 77 | 8.2 | 446 | US-09-761-716-2 | Sequence 2, Appl1 |
| 39 | 76.5 | 8.2 | 288 | US-08-875-062-1 | Sequence 1, Appl1 |
| 40 | 76.5 | 8.2 | 288 | US-09-007-288E-92 | Sequence 92, Appl1 |
| 41 | 76.5 | 8.2 | 933 | US-09-141-206-2 | Sequence 2, Appl1 |
| 42 | 76.5 | 8.2 | 933 | US-09-107-532A-6263 | Sequence 17, Appl |
| 43 | 76 | 8.1 | 323 | US-09-107-532A-6263 | Sequence 6263, Ap |
| 44 | 75.5 | 8.1 | 315 | US-09-252-991A-26131 | Sequence 26131, A |
| 45 | 75.5 | 8.1 | 325 | US-09-328-352-6196 | Sequence 6196, Ap |

ALIGNMENTS

RESULT 1
US-09-111-556A-3
Sequence 3, Application US/09111556A
Patent No. 6020180
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Parhar, Shankam A
APPLICANT: Egel-Mitani, Michi
APPLICANT: Borch, Kim
APPLICANT: Clausen, Ib G
APPLICANT: Hansen, Mogens T
TITLE OF INVENTION: C. ANTARCTICA LIPASE AND LIPASE VARIANTS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6020180 No. 6020180disk of No. 6020180th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/111,556A
FILING DATE: 22-DEC-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK PCT/DK93/00225
FILING DATE: 03-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3748.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-111-556A-3
Query Match 10.4%; Score 97; DB 3; Length 277;
Best Local Similarity 24.6%; Pred. No. 0.015;

| | Matches | 51; Conservative | 31; Mismatches | 79; Indels | 46; Gaps | 11. |
|----|---------|--|----------------|------------|----------|-----|
| QY | 3 | ILIVHGL-----YHGLVWHPRLSHRHKGYGTQTITSTNSLAIIDDEA-----IFPR | 48 | | | |
| Db | 11 | ILVHLLPFPDRIGSHHYFPG-----IQALNNEGQ-ASVVPPIISANDNEARGDQLLQQ | 64 | | | |
| QY | 49 | LD--RSLTASPNALVGHSLGLVLYKRLIESRASCETLSHVAALGSPLOGASIVKIB- | 105 | | | |
| Db | 65 | INHRLRQVQAQRVNLNIGHSQGLT-ARVVAIALP--ELINSYVSVSQPHNGSLDRRL | 121 | | | |
| QY | 106 | -----QLGLVALGNLSAEFG--LKHEDDESYPPQKSSIGATTPLGRSLILRDPLSDSG | 158 | | | |
| Db | 122 | AFVPGGLGETVAAALTTSPSAPILSLSGHPLPQNA-----LNLV---NALTTDG | 168 | | | |
| QY | 159 | TVTVEETKLAGMTDLHAISTSYENAV | 185 | | | |
| Db | 169 | VAAFNQYFQGLPDRKGGMGPAQVNAV | 195 | | | |

US-08-360-758-3
: Sequence 3, Application US/08360758
: Patent No. 6074863
: GENERAL INFORMATION:
: APPLICANT: Svendsen, Allan
: APPLICANT: Pathar, Shamkant A
: APPLICANT: Egel-Mitani, Michi
: APPLICANT: Borch, Kim
: APPLICANT: Clausen, Ib G
: APPLICANT: Hansen, Mogens T
: TITLE OF INVENTION: C. ANTARCTICA LIPASE AND LIPASE VARIANTS
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: No. 60748630 No. 6074863disk of No. 6074863th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Tape
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/360,758
: FILING DATE: 22-DEC-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK PCT/DK93/00225
: FILING DATE: 03-JUN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambiris, Elias J.
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 3748.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 277 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-360-758-3

```

Query Match      10.4% ; Score 97 ; DB 3 ; Length 277;
Best Local Similarity 24.6% ; Pred. No. 0.015;
Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;

QY      3 ILLVHGL-----YMHGLVMHPLSHRLAHKLYGRQLISNVSLAIDDEA---IFRR 48
|:||||| | ||| : |:| : : | | |:| : ::
```

[illegible]

RESULT 3
 US-09-252-991A-27337
 : Sequence 27337, Application US/09252991A
 : Patent No. 6551795
 : GENERAL INFORMATION:
 : APPLICANT: Marc J. Rubenfield et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS.
 : TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: 107196.136
 : CURRENT APPLICATION NUMBER: US/09/252,991A
 : CURRENT FILING DATE: 1999-02-18
 : PRIOR APPLICATION NUMBER: US 60/074,788
 : PRIOR FILING DATE: 1998-02-18
 : PRIOR APPLICATION NUMBER: US 60/094,190
 : PRIOR FILING DATE: 1998-07-27
 : NUMBER OF SEQ ID NOS: 33142
 : SEQ ID NO 27337
 : LENGTH: 351
 : TYPE: PRT
 : ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27337

[illegible]

```

RESULT 4
US-09-328-352-5069
; Sequence 5069, Application US/09328352
; Patent No. 6562938
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5069
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5069

```

```

Query Match      10.3%; Score 96.5; DB 4; Length 453;
Best Local Similarity 29.5%; Pred. No. 0.034;
Matches 39; Conservative 18; Mismatches 42; Indels 33; Gaps 7;

QY      3 IILVHGELYM-----GLVNPPLSHRLHK---LGYRT-QTISYNSLAIDDEAIFR 47

```

DB 164 VILVHGLMNLITWSNHYGIGERLLAQRHNMVLVNTVTRISANRS-----FS 217
QY 48 RLDSLTASNA---LVHSLGLVTKRYL---ESRAPSCTTSHVVALGSPLOGAS 99
DB 218 NLDLVLDRNPRITSIDLIGHSMGLVRSALFYKQNMVQIMHVENLVICISPHRGA- 276
QY 100 IVNKIEQLGLGV 111
DB 277 ---VIERGFAL 285

RESULT 5
US-09-107-532A-4594
Sequence 4594, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street
CITY: Waltham

STATE: Massachusetts
COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC
OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4594:

SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:
NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...316

SEQUENCE DESCRIPTION: SEQ ID NO: 4594:

US-09-107-532A-4594

Query Match 10.3%; Score 96; DB 4; Length 316;
Best Local Similarity 31.2%; Pred. No. 0.023; Mismatches 32; Indels 28; Gaps 6;
Matches 35; Conservative 17;

QY 61 LVHSLGLVTKRYL---ESRAPSCTTSHVVALGSPLOGAS 112
DB 170 LVHSLGLVTKRYL---ESRAPSCTTSHVVALGSPLOGAS 112
QY 113 LGNSAEFLGKEHDESRYPQKSGSIAGTIPLGRLSLLRDPLD---SDGTIV 160

DB 227 VVS-----NRY-QDYQWIGNIPITRREFLLAGQDERDLSGDTV 265

RESULT 6
US-08-034-650-10

Sequence 10, Application US/08034650
Patent No. 5641671

GENERAL INFORMATION:
APPLICANT: BOS, Janette W.

APPLICANT: BRENNEN, Leon G.

APPLICANT: VERBERS, Cornelis T.

APPLICANT: VISSER, Christiaan

TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE

TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

STREET: 1615 L. Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036-5601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/034,650

FILING DATE: 03-JUL-1991

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/727,235

FILING DATE: 03-JUL-1991

ATTORNEY/AGENT INFORMATION:
NAME: Kokulis, Paul N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: PNR/5970/91731

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-034-650-10

Query Match 10.3%; Score 96; DB 1; Length 358;
Best Local Similarity 21.2%; Pred. No. 0.027; Mismatches 65; Indels 62; Gaps 8;
Matches 43; Conservative 33;

QY 3 IIVHGL-----YHGLVMEPLSHRLKLYRTTISYNSLAID-----EAI 45
DB 50 VILVHGLAGTDKFRANVVDYWGIGSDLSH-----GAKYVYANLSGFQSDGPNRGEQL 104

QY 46 FRRLDRSL--THASPNALVHSLGLVTKRYLESRAPSCCTTSHVVALGSPLOGASIVUK 103
DB 105 LAYKQVLAATGATKYNLIGHSGQGLT--SRVYAAVAF--QVYASVTTTTPRGSSEFADF 161

QY 104 IEQLGLGVALLGNSAEFLGKEHDESRYPQKSGSIAGTIPLGRLSLLRDPLSDGTIV--- 160
DB 162 VQ-----DVLKTDPTGLSSTVIAAFVAVFGTLVSS 191

QY 161 --TVEETKLAGMTDHTAISTTSY 181
DB 192 SHNTQDALAALRTLTAAQTATY 214

RESULT 7
US-08-449-015-10

Sequence 10, Application US/08449015
Patent No. 5804409
GENERAL INFORMATION:
APPLICANT: BOS, Janette W.
APPLICANT: FRENKEN, Leon G.
APPLICANT: VERKIPS, Cornelis T.
APPLICANT: VISSER, Christiaan
TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE
TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARY & CUSHMAN
STREET: 1615 L. Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,015
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/727,235
FILING DATE: 03-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kokulis, Paul N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: PNK/5970/91731
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-449-015-10

Query Match 10.3%; Score 96; DB 1; Length 358;
Best Local Similarity 21.2%; Pred. No. 0.027;
Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;

QY 3 IILVHGL-----YMGGLVWHPLSRLKLGRTQTISVNSLAID-----EAI 45
DB 50 VILVHGLAGTDKFRANVVDWYWGQISDLQSH-----GAKYVYANLSEFOSDDGPNRGSQL 104
QY 46 FRRLDLSL--THASPNALVSHLGLVIRKYLESRAPSCETLSHVVAIGSPLOGASIVNK 103
DB 105 LAYVKQVLATGATKYNLIGSGGGLT--SRVAAVAP--QVASTYITGTPHRSGEFADF 161
QY 104 IEQLGLVALGNSAEFGLKEHDESRYPQKSGSIAGTIPLGLRSLLDPLDSDGTV--- 160
DB 162 VQ-----DVAKTDPTGLSGTVIAFAFVVFGLVSS 191
QY 161 --TWESTKTAGMTDHAISTSY 181
DB 192 SHNTDQDALALFRLTLTAQATY 214

RESULT 8
US-09-543-681A-6874
Sequence 6874, Application US/09543681A
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 6874
LENGTH: 333
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-6874

Query Match 9.9%; Score 92.5; DB 4; Length 333;
Best Local Similarity 23.8%; Pred. No. 0.06;
Matches 49; Conservative 27; Mismatches 67; Indels 63; Gaps 11;

QY 2 QIILVHGL-----YMGGL-----VMEPLSHRLKLGIR-----TQTSY 35
DB 68 RLVIFFGLGNFNSPYHGLAAAKARGLGVMM-----FRGSGEPNRQKRIY 117
QY 36 NSLAIDDEAIFRRRLDLSLTHASPNALVSHLGLVIRKYLESRAPSCETLSHVVAIGSP 95
DB 118 HSETEEDARYFLNMLKRGEOPTAIVGYSLSGNMLAYTLASGERA-VLDAAVIYSAPL 176
QY 96 QGASIVNKIEQLGLVALGNSAEF-----GLKEHDES--RYPQKSGSIAGTIPLGLRS 147
DB 177 MLEPCSTKIER-----GFSRFYQWYLLKGLKRNTRKLIHYPC-----SLPTSLLT 222
QY 148 L-LRDPDSDGTVYVEETKTAGMTD 172
DB 223 IKSIKRLRQPDILIT--AKIHGFKD 245

RESULT 9
US-08-602-359A-38
Sequence 38, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAPPIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-602-359A-38

Query Match
Best Local Similarity 24.6%; Pred. No. 0.054;
Matches 46; Conservative 16; Mismatches 58; Indels 67; Gaps 7;

QY 3 IIVHGLVMEGLVMEPLSHRLKLG-YRTQTSYNSLAIDDAIF 46
DB 16 VLVHGLGHSRGYRLIKELNYAGFYITDMPGHSKPKRGHTSVEEMELIDIE 75
QY 47 RLRLSLTHASPNALVHSLGLVTKRYLESR-----APSCETLSHYVAIG 92
DB 76 EIREAPF-----LFGSLGLVIRYAETRPDKIRGLIASSPALKSPETTPGMVAL- 127
QY 93 SPLGASIVNKIEQGLGVALGN--SAEPGKHEHDESRYPQKSGSIAGTIPGLRSL 150
DB 128 -----AKFLGKIAP---GVVLNGIKPFLSLRNDAVARY-----V 160

QY 151 RDPPLDSD 157
DB 161 EDPPLVHD 167

RESULT 10
US-08-732-412-2
Sequence 2, Application US/08732412

Patent No. 6017866
GENERAL INFORMATION:
APPLICANT: Ahle, Wolfgang
APPLICANT: Gerltse, Gjsbert
APPLICANT: Lenting, Hermann
TITLE OF INVENTION: LIPASES WITH IMPROVED
TITLE OF INVENTION: SURFACTANT RESISTANCE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: Genecor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/732,412
FILING DATE: 22-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/01687
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201253.5
FILING DATE: 04-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher L.
REGISTRATION NUMBER: 35,696
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-846-6504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-732-412-2

Query Match
Best Local Similarity 22.2%; Pred. No. 0.08;
Matches 51; Conservative 35; Mismatches 64; Indels 80; Gaps 12;

QY 3 IIVHGL-----YHGLVMEPLSHRLKLG-YRTQTSYNSLAIDDAIFRRDR 51
DB 39 IYVTHGLGFDLSLGVYWGJ---PSSLRSDGASVITVEQSLTSELRGELLEQVEE 95
QY 52 --SLTHASPNALVHSLGLVTKRYLESAPSCETLSHYVAIGSPLGASIVNKIEQL-- 107
DB 96 IAIISGKAVNLVGHSHGPTV-RYVAAYRP--DIVASTSYGAPRKSQSDTADFIQIRP 152
QY 108 -----GLGVALG-----NSAEFLKHEHDE-----SRYPQ----- 132
DB 153 GSAGEAIVAGIVNGALINFLSGSSSTSPQNALGLSELNSEGAAAFNAKYPQGIPTSA 212
QY 133 -----KSGSIAGTIP-----LGRSLLRDPDSDGTV 160
DB 213 CGEGAYKVGVSYSWSGTSPLTNVLDSDLLLGASSLTFDEP--NDGLV 260

RESULT 11
US-08-978-589A-2
Sequence 2, Application US/08978589A

Patent No. 6087145
GENERAL INFORMATION:
APPLICANT: ISHII, Takeshi
APPLICANT: MITSUDA, Satoshi
TITLE OF INVENTION: ESTERASE GENE AND ITS USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22040
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,589A
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 20-4336P
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-978-589A-2

Query Match
Best Local Similarity 26.9%; Pred. No. 0.11;
Matches 47; Conservative 20; Mismatches 57; Indels 51; Gaps 8;

QY 3 IIVHGL-----YHGLVME-----PLSHRLKLGIRYQTTSYNSLAIDDD-----EAIFFR 48
DB 55 IIVHGLGTGDKYGVVEYWRIFEDLRAH--GAAYVANTLSGFQSDGNGRGEDLLAF 112

QY 49 LDRSL--THASPNALVGHSLGGLVIRKYLESPAPSCETLSHVVAIGSPLOGASIVNKIEQ 106
 113 VKQVLAATGAQKVNLIHSGQGLT--SRVYASVAF--ELVASTTTISTPBMGSGFADPVOQ 169
 QY 107 LGLGVALGNSAFGLKEHDESRYPQKSGSIAGTTPGLGRSLILRDLPLDSGTVT 161
 170 L-----LQTDPTGLSTVLAGAFANALGTLT 194
 Db

RESULT 12
 US-09-336-601-1
 ; Sequence 1, Application US/09336601
 ; Patent No. 6184008
 ; GENERAL INFORMATION:
 ; APPLICANT: OHTA, Hiromichi
 ; APPLICANT: SUGAI, Takeshi
 ; APPLICANT: ISHII, Takeshi
 ; APPLICANT: MITSUDA, Satoshi
 ; TITLE OF INVENTION: PRODUCTION OF OPTICALLY ACTIVE SPHINGOID COMPOUND
 ; FILE REFERENCE: 2185-349P
 ; CURRENT APPLICATION NUMBER: US/09/336,601
 ; EARLIER FILING DATE: 1999-06-21
 ; EARLIER APPLICATION NUMBER: 09/034,007
 ; EARLIER FILING DATE: 1998-03-03
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 363
 ; TYPE: PRF
 ; ORGANISM: E. coli JM 109/pAL 612 strain
 US-09-336-601-1

Query Match
 Best Local Similarity 26.9%; Score 90.5; DB 3; Length 363;
 Matches 47; Conservative 20; Mismatches 57; Indels 51; Gaps 8;

QY 3 IIVHGL-----YHGLVNH-----PLSHRLKLGRTQTISVSLAID-----EAIPIR 48
 55 IIVHGLTGTDKXGVEEYWRIPEDLRH--GAAYVANLSGFOSDDGPNRGSQLAF 112
 Db 49 LDRSL--THASPNALVGHSLGGLVIRKYLESPAPSCETLSHVVAIGSPLOGASIVNKIEQ 106
 113 VKQVLAATGAQKVNLIHSGQGLT--SRVYASVAF--ELVASTTTISTPBMGSGFADPVOQ 169
 QY 107 LGLGVALGNSAFGLKEHDESRYPQKSGSIAGTTPGLGRSLILRDLPLDSGTVT 161
 170 L-----LQTDPTGLSTVLAGAFANALGTLT 194
 Db

RESULT 13
 US-09-219-120-2
 ; Sequence 2, Application US/09219120
 ; Patent No. 6472188
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKESHI, Ishii
 ; APPLICANT: SATOSHI, Mitsuuda
 ; TITLE OF INVENTION: ESTERASE GENE AND ITS USE
 ; FILE REFERENCE: 20-4336P
 ; CURRENT APPLICATION NUMBER: US/09/219,120
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 363
 ; TYPE: PRF
 ; ORGANISM: Burkholderia cepacia
 US-09-219-120-2

Query Match
 Best Local Similarity 26.9%; Score 90.5; DB 4; Length 363;
 Matches 47; Conservative 20; Mismatches 57; Indels 51; Gaps 8;

QY 3 IIVHGL-----YHGLVNH-----PLSHRLKLGRTQTISVSLAID-----EAIPIR 48
 55 IIVHGLTGTDKXGVEEYWRIPEDLRH--GAAYVANLSGFOSDDGPNRGSQLAF 112
 Db 49 LDRSL--THASPNALVGHSLGGLVIRKYLESPAPSCETLSHVVAIGSPLOGASIVNKIEQ 106
 113 VKQVLAATGAQKVNLIHSGQGLT--SRVYASVAF--ELVASTTTISTPBMGSGFADPVOQ 169
 QY 107 LGLGVALGNSAFGLKEHDESRYPQKSGSIAGTTPGLGRSLILRDLPLDSGTVT 161
 170 L-----LQTDPTGLSTVLAGAFANALGTLT 194
 Db

RESULT 14
 US-09-252-991A-27871
 ; Sequence 27871, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196,136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; EARLIER FILING DATE: 1999-02-18
 ; EARLIER APPLICATION NUMBER: US 60/074,788
 ; EARLIER FILING DATE: 1998-02-18
 ; EARLIER APPLICATION NUMBER: US 60/094,190
 ; EARLIER FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 27871
 ; LENGTH: 344
 ; TYPE: PRF
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27871

Query Match
 Best Local Similarity 24.5%; Score 89; DB 4; Length 344;
 Matches 50; Conservative 28; Mismatches 86; Indels 40; Gaps 7;

QY 9 IYHGLVNHPLSH-----RLHKLGVRTQTISVSLAID-----EAIPIR 47
 75 LAHGLTGSSSSHYILGLQALBERGMASVALNWRGSGEPNRLPRGYSGVSDNLAEVV 134
 Db 48 LDRSL--THASPNALVGHSLGGLVIRKYLESPAPSCETLSHVVAIGSPLOGASIVNKIEQ 107
 135 AHLRARPQAPLYAVGVSLGAVNLKYLGETAGDCEPLGG--VAVSVFFR--LDECADRI 190
 QY 108 LGLGVALGNSAFGLKEHDESRYPQKSGSIAGTTPGLGRSLILRDLPLDSGTVT 157
 191 GLGFSRVYQAHFKAMLAIVYQDKRLFGEGQGLEALQRLGRLPBGKRTF-----WDFD 245
 Db 158 GTVVEETKLAGMTDHAISTSY 181
 246 GRVYAPLHGFPADARDVYRASSRY 269
 QY

RESULT 15
 US-08-637-759B-349
 ; Sequence 349, Application US/08637759B
 ; Patent No. 5876931
 ; GENERAL INFORMATION:
 ; APPLICANT: David William Holden
 ; TITLE OF INVENTION: Identification of Genes
 ; NUMBER OF SEQUENCES: 501
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Patrice L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: REWS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 349:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-349

Query Match 9.5%; Score 88.5; DB 2; Length 151;
Best Local Similarity 21.8%; Pred. No. 0.052;
Matches 37; Conservative 27; Mismatches 77; Indels 29; Gaps 6;

QY 6 VHGVTMGVWVPELSHRKLGRTQTSYNSLAIDDEAFRRIDRSITHTSPNALVGH 65
DB 3 VFNVRKHGDKRHPIDRRSRNPAADTAERFRTYMAIDKNIVHRNTHQAKSHHTFRFG 62
QY 66 LGGLVTKRYLESRAPECTLSHVAIGSPLOGASIVNK-IEOLGVALGNSAEFG 121
DB 63 QTFALVSRYLKEKV-SC-----APQRAKITHGFIGQRINIMHRADNVSGI 111
QY 122 --KEHDDSRYPQKSGSIAGTIFLGLRSLLRPLDSQIVTVVEETKIAG 169
DB 112 DHQHGDYARQPE-----PL---SNLWEDTLTTAGAIELRNHRQ 149

Search completed: April 29, 2004, 11:37:49
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:37:05 ; Search time 43 Seconds
(without alignments)
1192.560 Million cell updates/sec

Title: US-10-603-260-5
Sequence: 1 MQLIVHGLVYMGVWHPPLS.....KINAMTDHIAISTTSENAV 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues
Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/2/pubppaa/US05_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/2/pubppaa/US04_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubppaa/US03_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubppaa/US02_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubppaa/US01_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubppaa/US00_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubppaa/US05_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubppaa/US04_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubppaa/US03_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubppaa/US02_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubppaa/US01_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubppaa/US00_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1 | 936 | 100.0 | 185 | US-10-603-260-5 | Sequence 5, Appl1 |
| 2 | 110 | 11.8 | 249 | US-10-369-493-8213 | Sequence 8213, Ap |
| 3 | 97 | 10.4 | 447 | US-10-421-654-16 | Sequence 16, Appl |
| 4 | 91.5 | 9.8 | 262 | US-09-903-410-38 | Sequence 38, Appl |
| 5 | 91.5 | 9.8 | 262 | US-10-027-805-38 | Sequence 38, Appl |
| 6 | 91.5 | 9.8 | 262 | US-10-027-804-38 | Sequence 38, Appl |
| 7 | 90.5 | 9.7 | 286 | US-10-156-761-10998 | Sequence 10998, A |
| 8 | 90.5 | 9.7 | 286 | US-10-147-467-2 | Sequence 2, Appl1 |
| 9 | 90 | 9.6 | 443 | US-10-434-599-24654 | Sequence 24654, A |
| 10 | 88.5 | 9.5 | 518 | US-10-038-174-1 | Sequence 1, Appl1 |
| 11 | 88.5 | 9.5 | 627 | US-10-282-122A-70475 | Sequence 70475, A |
| 12 | 87.5 | 9.3 | 286 | US-10-156-761-14086 | Sequence 14086, A |
| 13 | 87.5 | 9.3 | 345 | US-10-282-122A-47438 | Sequence 47438, A |
| 14 | 87 | 9.3 | 345 | US-10-282-122A-76642 | Sequence 76642, A |
| 15 | 87 | 9.3 | 505 | US-09-938-803-20 | Sequence 20, Appl |

| | | | | | |
|----|------|-----|------|----------------------|--------------------|
| 16 | 85 | 9.1 | 1216 | US-10-369-493-1505 | Sequence 1505, Ap |
| 17 | 82.5 | 8.8 | 275 | US-10-369-493-4463 | Sequence 4463, Ap |
| 18 | 82.5 | 8.8 | 275 | US-10-369-493-7223 | Sequence 7223, Ap |
| 19 | 82.5 | 8.8 | 279 | US-10-282-122A-62627 | Sequence 62627, A |
| 20 | 82.5 | 8.8 | 301 | US-10-282-122A-49252 | Sequence 49252, A |
| 21 | 82.5 | 8.8 | 323 | US-10-282-122A-64350 | Sequence 64350, A |
| 22 | 82.5 | 8.8 | 432 | US-10-156-761-13452 | Sequence 13452, A |
| 23 | 82.5 | 8.8 | 470 | US-10-421-654-58 | Sequence 58, Appl |
| 24 | 82 | 8.8 | 244 | US-10-289-762-560 | Sequence 360, Appl |
| 25 | 81.5 | 8.7 | 359 | US-10-369-493-21672 | Sequence 21672, A |
| 26 | 81 | 8.7 | 332 | US-10-050-882-99 | Sequence 99, Appl1 |
| 27 | 80.5 | 8.6 | 244 | US-10-369-493-19863 | Sequence 19863, Ap |
| 28 | 80.5 | 8.6 | 252 | US-10-369-493-16706 | Sequence 16706, A |
| 29 | 80.5 | 8.6 | 259 | US-10-369-493-16706 | Sequence 16706, A |
| 30 | 80.5 | 8.6 | 267 | US-10-282-122A-46834 | Sequence 46834, A |
| 31 | 80.5 | 8.6 | 301 | US-10-156-761-7846 | Sequence 7846, Ap |
| 32 | 80.5 | 8.6 | 331 | US-09-815-242-10799 | Sequence 10799, A |
| 33 | 80.5 | 8.6 | 331 | US-10-282-122A-57052 | Sequence 57052, A |
| 34 | 80.5 | 8.6 | 360 | US-10-424-599-221830 | Sequence 221830, A |
| 35 | 80.5 | 8.6 | 361 | US-09-896-908-2 | Sequence 2, Appl1 |
| 36 | 80.5 | 8.6 | 1460 | US-10-282-122A-55255 | Sequence 55255, A |
| 37 | 80 | 8.5 | 225 | US-10-425-114-63517 | Sequence 63517, A |
| 38 | 80 | 8.5 | 240 | US-10-425-114-65130 | Sequence 45130, A |
| 39 | 80 | 8.5 | 257 | US-10-282-122A-77391 | Sequence 77391, A |
| 40 | 80 | 8.5 | 587 | US-10-425-114-68123 | Sequence 68123, A |
| 41 | 80 | 8.5 | 1129 | US-10-156-761-8946 | Sequence 8946, Ap |
| 42 | 79.5 | 8.5 | 302 | US-10-282-122A-47335 | Sequence 47335, A |
| 43 | 79.5 | 8.5 | 303 | US-10-282-122A-50776 | Sequence 50776, A |
| 44 | 79.5 | 8.5 | 403 | US-10-156-761-14428 | Sequence 14428, A |
| 45 | 79 | 8.4 | 270 | US-10-206-576-188 | Sequence 188, Appl |

ALIGNMENTS

RESULT 1
US-10-603-260-5
; Sequence 5, Application US/10603260
; Publication No. US20040009570A1
; GENERAL INFORMATION:
; APPLICANT: CJ Corporation
; TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a
; TITLE OF INVENTION: nucleotide sequence encoding the same
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/603,260
; CURRENT FILING DATE: 2003-06-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Kopatentim 1.71
; SEQ ID NO 5
; LENGTH: 185
; TYPE: PRT
; ORGANISM: *Vibrio metchnikovii* RH530
US-10-603-260-5

| | | | | |
|-----------------------|-----------------|---|-----------|-------------|
| Query Match | 100.0% | Score 936; | DB 15; | Length 185; |
| Best Local Similarity | 100.0% | Pred. No. 4.1e-96; | | |
| Matches 185; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| QY | 1 | MQLIVHGLVYMGVWHPPLSRLHLKLGRTQTISTNSLAIDDEAIFRRDRLSTASPA | 60 | |
| DB | 1 | MQLIVHGLVYMGVWHPPLSRLHLKLGRTQTISTNSLAIDDEAIFRRDRLSTASPA | 60 | |
| QY | 61 | LVHSGIGLVYIKRYLESRAAPSCETLSHVVAIGSPLOGASIVNKIKRLGIGVALGNSAERG | 120 | |
| DB | 61 | LVHSGIGLVYIKRYLESRAAPSCETLSHVVAIGSPLOGASIVNKIKRLGIGVALGNSAERG | 120 | |
| QY | 121 | LKHHDESRYPKSGSIAGTIPGLGRSLILRPLDSDGTVTVETKIAQMTDHAISTTS | 180 | |
| DB | 121 | LKHHDESRYPKSGSIAGTIPGLGRSLILRPLDSDGTVTVETKIAQMTDHAISTTS | 180 | |
| QY | 181 | YENAV 185 | | |
| DB | 181 | YENAV 185 | | |

APPLICANT: ROBERTSON, Daniel E.
MURPHY, Dennis
REID, John
MAFFIA, Anthony
LINK, Steven
SWANSON, Ronald V.
WARREN, Patrick V.
KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,805
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-10-027-805-38
Query Match
Best Local Similarity 9.8%; Score 91.5; DB 13; Length 262;
Matches 46; Conservative 16; Mismatches 58; Indels 67; Gaps 7;
QY 3 IILVHGIVMGGLVWHPPLSHRLHKLGYRTQTISY-----NSLAIDDEAIF 46
DB 16 VLVVHGLGHSRGYGRILIKELNYAGFVYTFDWPFGHKSPGKRGHTSVEMAEIIDSIIIE 75
QY 47 RLDRSLTHASPNALVGHSLGGLVIRYLESR-----APSCETLSHVAIG 92
DB 76 EIREKRF-----LFGSLGGLTVIRYAEIRPDKIRGLIASPALAKSPETGFWAL- 127
QY 93 SPLQASIVNKIKIQGLGVALGN--SAEFGLEKHDDSESRYPQSSGSIAGTIPLGLSILL 150
DB 128 -----AKPIGKIAP---GVVLSNGIKPELISRNRDAVRY-----V 160
QY 151 RDPPLSD 157
DB 161 EDPPLVHD 167
RESULT 5
US-10-027-804-38
Sequence 38, Application US/10027804
Publication No. US20030054530A1
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
MURPHY, Dennis

REID, John
MAFFIA, Anthony
LINK, Steven
SWANSON, Ronald V.
WARREN, Patrick V.
KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,804
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-10-027-804-38
Query Match
Best Local Similarity 9.8%; Score 91.5; DB 14; Length 262;
Matches 46; Conservative 16; Mismatches 58; Indels 67; Gaps 7;
QY 3 IILVHGIVMGGLVWHPPLSHRLHKLGYRTQTISY-----NSLAIDDEAIF 46
DB 16 VLVVHGLGHSRGYGRILIKELNYAGFVYTFDWPFGHKSPGKRGHTSVEMAEIIDSIIIE 75
QY 47 RLDRSLTHASPNALVGHSLGGLVIRYLESR-----APSCETLSHVAIG 92
DB 76 EIREKRF-----LFGSLGGLTVIRYAEIRPDKIRGLIASPALAKSPETGFWAL- 127
QY 93 SPLQASIVNKIKIQGLGVALGN--SAEFGLEKHDDSESRYPQSSGSIAGTIPLGLSILL 150
DB 128 -----AKPIGKIAP---GVVLSNGIKPELISRNRDAVRY-----V 160
QY 151 RDPPLSD 157
DB 161 EDPPLVHD 167
RESULT 7
US-10-156-761-10998
Sequence 10998, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMTA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI

Db 107 VVFWHGFV--GLVGEDAFSMYPNGTKYNVKOELFTLGYRVEANVGAFFSSNYDRAVE 164
 QY 46 -----FRRLDRSLTHASPN-----LVGSHLGLVYK--RY 74
 Db 165 LYYIKGGRVVDYGAHAHAKYGHKRYGRTYEGIMPDMEBKHHLVGSHWGQTLRLMEHF 224
 QY 75 LES-----RAPSCETLSHVVAIGSPLOGA-----SIV 101
 Db 225 LRNGGEIDYORQYGVSDLFKGGQDNMVTITTLGTPHNGTPAADKLGSTFKYDIT 264
 QY 102 NKIEQLG---LGVALGNSAEFGLKEHDESRYPQKSGSIAGT 140
 Db 285 NRIGKIGTKALDLELGFSS-OWGFKQCPNES-YAEVAKRIANS 325

RESULT 11

US-10-282-122A-70475
 ; Sequence 70475, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 70475
 ; LENGTH: 627
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-10-282-122A-70475

Query Match 9.5%; Score 88.5; DB 12; Length 627;
 Best Local Similarity 21.1%; Pred. No. 1.1;

Matches 47; Conservative 32; Mismatches 55; Indels 89; Gaps 12;

QY 3 IIVHGLYMHGIV-----MHP-----LSHRHLKLGRTYRTISYNSIAID-DEAI- 45
 Db 248 VVFWHGFV--GLVGEDAFSMYPNGTKYNVKOELFTLGYRVEANVGAFFSSNYDRAVE 305

QY 46 -----FRRLDRSLTHASPN-----LVGSHLGLVYK--RY 74
 Db 306 LYYIKGGRVVDYGAHAHAKYGHKRYGRTYEGIMPDMEBKHHLVGSHWGQTLRLMEHF 365
 QY 75 LES-----RAPSCETLSHVVAIGSPLOGA-----SIV 101
 Db 366 LRNGGEIDYORQYGVSDLFKGGQDNMVTITTLGTPHNGTPAADKLGSTFKYDIT 425
 QY 102 NKIEQLG---LGVALGNSAEFGLKEHDESRYPQKSGSIAGT 140
 Db 426 NRIGKIGTKALDLELGFSS-OWGFKQCPNES-YAEVAKRIANS 466

RESULT 12

US-10-156-761-14086
 ; Sequence 14086, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 14086
 ; LENGTH: 286
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-14086

Query Match 9.3%; Score 87.5; DB 14; Length 286;
 Best Local Similarity 23.5%; Pred. No. 0.45;
 Matches 53; Conservative 23; Mismatches 53; Indels 97; Gaps 12;

QY 3 IIVHGLYMHGIV-----GLVHAPLSRHLKLGRTYRTISYNSIA-----IDEA- 44
 Db 53 VVFWHGTGNSVDNMLGLAPY-LEHR-----GYCVFSLDYGQSGVFFHGLGPIIDKSAQ 107
 QY 45 --IFRLDRSL--THASPNALVGHSLGIVIRKYLESRAPSCETLSHVVAIGSPLOGASI 100
 Db 108 LQVF--VDKVLTPATGATKADLVGHSQGMPRYTLKF-----LGAGAK 148
 QY 101 VNKIEQLGVALGNSAEFGLKEHDESRYPQKSGSIAGTIPLGRLSLRLPDLSDGTV 160
 Db 149 VNAL-----VGIAPNNH-----CTTISGLTNLLPYFP--GAE 178
 QY 161 TVEETKIAKMTD-----HIAISTSYEMAY 185
 Db 179 DLSTATPGIADQVVGSAFMAKLINAGDTVAGVHTVATATQYDEVV 224

RESULT 13

US-10-282-122A-47436
 ; Sequence 47436, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel

```

; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47438
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47438

Query Match
Best Local Similarity 9.3%; Score 87.5; DB 12; Length 364;
Matches 34; Conservative 20; Mismatches 43; Indels 27; Gaps 6;

QY 3 IIVHGL-----YHGLVHMPPLSHRLHKGYRQTSYNSLAID-----DAI 45
DB 55 IIVHGLTGTGDXAGVLDWYTG-----IQEDLQGHATVYVNLGFSDDSPNGRGHQL 109
QY 46 FRRLDRSL--THASPNALVGHSLGSLVIRKYLESRAPSCETSHVAIGSPLOGASIVNK 103
DB 110 LAYVKTIVLANTGATKYNLVGHSGGLT--SRVAAVAP--DIVASVTTTCTPHRGSEFPDF 166

QY 104 IEQL 107
DB 167 VQGV 170

RESULT 14
US-10-282-122A-76642
; Sequence 76642, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```

; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76642
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Treponema pallidum
US-10-282-122A-76642

Query Match
Best Local Similarity 9.3%; Score 87; DB 12; Length 345;
Matches 37; Conservative 26; Mismatches 66; Indels 12; Gaps 4;

QY 45 IFRRLDRSLTHASPNALVGHSLGSLVIRKYLESRAPSCETSHVAIGSPLOGASIVN-- 102
DB 154 LYRVDFTRV-----GLVHSGWGTALAHY--ANKYP--ERIRTVVLAHPNGADWVNMV 205
QY 103 --KIEQLGLGVALGNLSAFGLKEDDSRRYPQXGSIAGTIPGLRSLLRDPLSDGTV 160
DB 206 GGNKRGQGTDLRSVRAVLLITGKQMLAYPERSYEPFQGLPDVDVACFLSKMGHGAG 265

QY 161 TVEETKLAGMTDHAISTSY 181
DB 266 TDKYFVWGFTYKTLVSESF 286

RESULT 15
US-09-938-803-20
; Sequence 20, Application US/09938803
; Patent No. US20020076762A1
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Reddy, Roopa
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Yang, Junming
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: FULL-LENGTH EXPRESSED GENETIC MARKERS
; FILE REFERENCE: PP-0695 US
; CURRENT APPLICATION NUMBER: US/09/938,803
; CURRENT FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/311,894
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone 3039890
US-09-938-803-20

```

```

Query Match          9.3%; Score 87; DB 9; Length 505;
Best Local Similarity 27.0%; Pred. No. 1.2;
Matches 37; Conservative 20; Mismatches 54; Indels 26; Gaps 6;

```

```

Qy 49 LDRSLTHASPVNLYGHSIGLVIRYDESRAPSCETLSHYVAIGSPLOGASIVNKIEQL-107
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 42 LKLMFEEELRNACRGQGVGF-----PAMKQIGNVAAL-----PGIVRSIGLP 86

Qy 108 ----GLGVALGNLSAEFGKHEHDESRYPQKSGSIAGTIPGLRSLLRDPLDSDGTVYE 163
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 87 DVHSGYGFALIGMAAFDY--NDPEA--VSRFGVGFDPINGVR--LIRTNLDESQVQPVK 140

Qy 164 ETKIAGMTDHTIAISTTS 180
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 141 EQLAQAMFDPHIPGVGS 157

```

```

Search completed: April 29, 2004, 11:42:30
Job time : 44 secs

```